

NONINFORMATIVE PRIORS AND BAYESIAN INFERENCE

By

YEONG-HWA KIM

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
UNIVERSITY OF FLORIDA

2000

Copyright 2000

by

Yeong-Hwa Kim

*Dedicated to God for giving me strength,
my parents,
my brother (Yeong-Taeg),
my wife (Mee-Jung),
my daughter (Ha-Na), and my son (Tae-Meen)
with all my love from Jesus.*

ACKNOWLEDGEMENTS

Like many things in life, obtaining a Ph.D. degree is a task that cannot be accomplished singlehandedly. I am grateful to all of those persons who were directly or indirectly responsible in the production of this dissertation.

My advisor, the Distinguished Professor Malay Ghosh, who served as the Ph.D. committee chair, is to be thanked for his guidance, mentoring, and cordial attention over the years. He lent a great deal of support and encouragement, and kept a watchful eye over me through the graduate program. Professors Ramon Littell, Beverly Brechner, James Booth and James Hobert, who compose the remainder of the Ph.D. committee, are also to be thanked for their expert advice and guidance. I also would like to give my special thanks to Dr. Ronald Randles, Chairman of the Department of Statistics, who supported me financially throughout my years at the University of Florida.

I express heartfelt gratitude to my parents, brother, and sister for their never-ending support, encouragement, prayer and love throughout all my endeavors. I will forever be grateful to them for their encouragement to continue my study and for the dreams they hold for me.

Finally, but most importantly, special thanks to my wife, Mee-Jung. I could not have completed this work without her unwavering love, patience, and prayer. Even when I did not, she always had confidence in the Lord, that I would complete this dissertation. I also would like to express my enormous love to my daughter, Ha-Na, and my son, Tae-Meen.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vi
LIST OF FIGURES	vii
ABSTRACT	viii
CHAPTERS	
1 INTRODUCTION	1
1.1 Literature Review	1
1.2 The Subject of This Dissertation	10
2 THE BEHRENS-FISHER PROBLEM	12
2.1 Introduction	12
2.2 Development of Noninformative Priors	16
2.2.1 Probability Matching Priors	16
2.2.2 HPD Matching Prior	20
2.2.3 Matching via Inversion of Posterior Bartlett Corrected Likelihood Ratio Statistics	21
2.2.4 Jeffreys' General Rule Prior	21
2.2.5 Reference Priors	22
2.3 Propriety of Posteriors and Simulation Study	26
2.4 Conditional Frequentist Properties	32
2.5 Concluding Remarks	33
3 INTERVAL ESTIMATION OF THE COMMON MEAN OF SEV- ERAL NORMAL POPULATIONS	35
3.1 Introduction	35
3.2 Noninformative Priors	37
3.2.1 Jeffreys' Prior	37
3.2.2 Reference Priors	38
3.2.3 Quantile Matching Priors	39
3.2.4 HPD Matching Prior	40

3.2.5	Matching via Inversion of Posterior Bartlett Corrected Likelihood Ratio Statistics	41
3.3	Propriety of Posteriors	41
3.4	Simulation Study	43
3.5	Comparison with Other Frequentist Methods	44
3.5.1	Confidence Interval for μ Based on T_i 's	46
3.5.2	Confidence Interval for μ Based on F_i 's	47
3.5.3	Confidence Regions for μ Based on P_i 's	48
3.5.4	Data Analysis	50
3.6	Concluding Remarks	53
4	BAYESIAN INFERENCE FOR THE RATIOS OF REGRESSION COEFFICIENTS IN A LINEAR MODEL	55
4.1	Introduction	55
4.2	Some Special Examples	57
4.2.1	The Linear Calibration Problem	57
4.2.2	The Fieller-Creasy Problem	58
4.2.3	Parallel-Line Bioassay Problem	59
4.2.4	Slope-Ratio Bioassay Problem	60
4.3	The Orthogonal Transformation	62
4.4	Development of Noninformative Priors	64
4.4.1	Jeffreys' Prior	64
4.4.2	Reference Priors	64
4.4.3	Probability Matching Priors	66
4.5	Propriety of Posteriors	68
4.6	Numerical Examples	72
4.7	Appendix	74
5	SUMMARY AND FUTURE RESEARCH	83
5.1	Summary	83
5.2	Future Research	84
	BIOGRAPHICAL SKETCH	93

LIST OF TABLES

<u>Table</u>	<u>page</u>
2.1 Catalog of reference priors for Behrens-Fisher Problem	25
2.2 Frequentist probabilities of 0.05 (0.95) posterior quantiles of θ_1 when $\mu_1 = 2.0, \mu_2 = 0.0$ and $\sigma_2^2 = 1.0$ ($n_1 > n_2$)	30
2.3 Frequentist coverage probabilities of 0.05 (0.95) posterior quantiles of θ_1 when $\mu_1 = 2.0, \mu_2 = 0.0$ and $\sigma_2^2 = 1.0$ ($n_1 < n_2$)	31
3.1 Frequentist coverage probabilities of 0.05 (0.95) posterior quantiles of μ when $\mu = 0.0$ and $\sigma_2^2 = 1.0$ ($n_1 > n_2$)	45
3.2 Frequentist coverage probabilities of 0.05 (0.95) posterior quantiles of μ when $\mu = 0.0$ and $\sigma_2^2 = 1.0$ ($n_1 < n_2$)	45
3.3 Percentage of albumin in plasma protein	51
3.4 Interval Estimates for μ	51
3.5 Selenium in non-fat milk powder	52
3.6 Interval Estimates for μ	53
4.1 Catalog of the elements of $\mathbf{C}$	71
4.2 Responses in an assay of vitamin B_{12}	72
4.3 Posterior quantiles	73
4.4 Responses in an assay of riboflavin in malt	73
4.5 Posterior quantiles	74

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
3.1 Posterior pdf of μ	52
3.2 Posterior pdf of μ	54

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

NONINFORMATIVE PRIORS AND BAYESIAN INFERENCE

By

Yeong-Hwa Kim

August 2000

Chairman: Malay Ghosh
Major Department: Statistics

The prime focus of this dissertation is Bayesian analysis of some classical problems of statistical inference. In particular, we have derived several “default” priors such as Jeffreys’ prior, reference priors, and probability matching priors for the Behrens-Fisher problem, estimation of the common mean of several normal populations, and for the ratios of regression coefficients in a linear model. The latter includes, as special cases, the linear calibration problem, the Fieller-Creasy problem, slope-ratio bioassay, and parallel-line bioassay.

For the Behrens-Fisher problem, we find an orthogonal transformation in the sense of Cox and Reid and propose an alternative prior such that the coverage probability of the resulting credible interval for the difference of the two means matches asymptotically the corresponding frequentist coverage probability more accurately than Jeffreys’ prior. Though the result holds only asymptotically, our simulation study indicates that second order probability matching prior performs very well in terms of matching the target coverage probabilities in a frequentist sense, especially

for small and moderate sample sizes. The prior is also justified from the conditional frequentist perspective.

For the interval estimation of the common mean of several normal populations, Jeffreys' prior, reference priors, a class of first order probability matching priors and a second order probability matching prior are derived. Though asymptotic, our simulation results indicate that the second order probability matching prior meets the frequentist target coverage probabilities better than reference prior especially for small and moderate sample sizes. Several confidence intervals that are proposed from a frequentist point of view are compared against the Bayesian method by analyzing two real examples. The lengths of the Bayesian intervals are nearly equivalent to the shortest frequentist intervals under both circumstances. Also, in this case, it turns out that among the frequentist procedures, there is no clear-cut winner.

For the ratios of regression coefficients in a linear model, we find an orthogonal transformation and derive Jeffreys' prior, reference priors and a class of first order probability matching priors. A second order probability matching prior is also derived and proprieties of posterior distributions are discussed. Several real examples are analyzed.

CHAPTER 1

INTRODUCTION

In this chapter we discuss some pertinent literature related to the development of “default” or “noninformative” priors. Also, we introduce the topics of this dissertation.

1.1 Literature Review

Bayesian techniques have found wider acceptance in recent years in the theory and practice of statistics. These can be partly explained by the fact that even in the presence of little or vague prior information, Bayesian techniques can often be used successfully by employing the so-called “diffuse”, “vague” or “noninformative” priors. Thus, not surprisingly, over the years, a wide range of noninformative priors has been proposed and studied.

When prior distributions have no population basis, they can be difficult to construct, and there has long been a desire for prior distributions that can be guaranteed to play a minimal role in the posterior distribution. Such priors are often called vague, diffuse or *noninformative* priors. The rationale for using noninformative priors is often said to be “let the data speak for themselves” so that inferences are unaffected by current prior.

The earliest use of noninformative priors is attributed to Laplace (1812) who recommended using a flat prior over the entire parameter space. However, a uniform prior lacks invariance under reparameterization since a uniform distribution under one parameterization will not yield, on transformation, another uniform distribution unless the transformation is linear. For example, a uniform prior for the standard

deviation σ will not transform into a uniform prior for the variance σ^2 . This lack of invariance of the uniform prior often translates into considerable variation in the resulting posteriors. To overcome this difficulty, Jeffreys (1961) proposed the prior proportional to the positive square root of the determinant of the Fisher information matrix, which remains invariant under any one-to-one parameterization. Despite its success in one-parameter problems, Jeffreys' prior often runs into serious difficulties in multiparameter problems when only a subset of one or more parametric function(s) of the parameter vector $\theta(p \times 1)$ is of inferential interest and the remaining parameters are nuisance parameters. For example, Berger and Bernardo (1992a) showed that Jeffreys' prior can lead to an inconsistent estimator of the error variance in the balanced one-way normal ANOVA model when the number of unknown treatment parameters goes to infinity in direct proportion to sample size. Here Jeffreys' prior fails to avoid the Neyman-Scott (1948) phenomenon.

To overcome these deficiencies of Jeffreys' prior, Bernardo (1979) introduced the so-called "reference prior" for multiparameter problems by dividing the parameter vector into parameters of interest and nuisance parameters. This idea was further extended and generalized by Berger and Bernardo (1989, 1992a,b) by splitting the parameter vector into two or more groups according to their order of importance. As pointed out by these authors, Jeffreys' prior is also a reference prior when all the parameters are treated as equally important.

The idea behind the development of a reference prior is as follows: For an experiment with density $p(\mathbf{x}|\theta)$ and prior density $\pi(\theta)$, consider the amount of information about θ that the experiment can be expected to provide. Writing $\pi(\theta|\mathbf{x})$ as the posterior of θ given $\mathbf{x}$, Bernardo (1979) argues for using, as a measure of this information, $I^\theta\{p, \pi\} = \int \int p(\mathbf{x}|\theta)\pi(\theta) \log \frac{\pi(\theta|\mathbf{x})}{\pi(\theta)} d\theta d\mathbf{x}$.

The reference prior is the π that maximizes this quantity, the rationale being that the larger this information is, the less informative is the prior. For a variety of technical reasons, the reference prior is actually defined, not for the experiment $p(\mathbf{x}|\boldsymbol{\theta})$, but via an asymptotic limit of *iid* repetitions of the experiment. In situations where asymptotic normality of the posterior holds, Bernardo (1979) showed that the reference prior for $\boldsymbol{\theta}$, provided there are no nuisance parameters, is Jeffreys' (1961) prior $\pi(\boldsymbol{\theta}) = |\mathbf{I}(\boldsymbol{\theta})|^{1/2}$. Here $\mathbf{I}(\boldsymbol{\theta})$ is the Fisher information matrix and $|\mathbf{A}|$ denotes the determinant of a matrix $\mathbf{A}$. Such, however, is not the case in the presence of nuisance parameters.

While Bernardo (1979) considered only two-group reference priors by splitting the parameters into two groups – parameters of interest and nuisance parameters – Berger and Bernardo (1992b) considered the more general case of m -group reference priors. We will follow their notations to introduce the algorithm of the general m -group reference priors in a particular set up which covers most examples that are available in the literature.

Consider a parametric statistical problem in which the random observation $\mathbf{X}$ has density $p(\mathbf{x}|\boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta \subset \mathcal{R}^k$ is the unknown parameter. We assume that the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta}) = -E_{\mathbf{x}|\boldsymbol{\theta}} \left[\left\{ \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log p(\mathbf{x}|\boldsymbol{\theta}) \right\} \right]$ exists and has rank k , so that it is invertible.

We assume that the θ_i are separated into m groups of sizes $n_1, \dots, n_m$, and that these groups are given by

$$\boldsymbol{\theta}_{(1)} = (\theta_1, \dots, \theta_{n_1}), \quad \boldsymbol{\theta}_{(2)} = (\theta_{n_1+1}, \dots, \theta_{n_1+n_2}), \dots,$$

$$\boldsymbol{\theta}_{(i)} = (\theta_{N_{i-1}+1}, \dots, \theta_{N_i}), \dots, \boldsymbol{\theta}_{(m)} = (\theta_{N_{m-1}+1}, \dots, \theta_k),$$

where $N_j = n_1 + \dots + n_j$ for $j = 1, \dots, m$. Also, we define, for $j = 1, \dots, m$,

$$\boldsymbol{\theta}_{[j]} = (\theta_{(1)}, \dots, \theta_{(j)}), \quad \boldsymbol{\theta}_{[\sim j]} = (\theta_{(j+1)}, \dots, \theta_{(m)})$$

with $\boldsymbol{\theta}_{[m]} = \boldsymbol{\theta}$ and $\boldsymbol{\theta}_{[\sim m]} = \phi$, the empty set. Here $\boldsymbol{\theta}_{(1)}$ is the parameter of interest. For the construction of reference priors, often the ordering of the remaining groups does not matter.

Let $\Theta^l \subset \Theta, l = 1, 2, \dots$, denote the compact subsets to be chosen for $\boldsymbol{\theta}$ such that

$$\bigcup_{l=1}^{\infty} \Theta^l = \Theta \quad .$$

For $j = 0, 1, \dots, m-1$, let

$$\Theta^l(\boldsymbol{\theta}_{[j]}) = \left\{ \boldsymbol{\theta}_{(j+1)} : (\boldsymbol{\theta}_{[j]}, \boldsymbol{\theta}_{(j+1)}, \boldsymbol{\theta}_{[\sim(j+1)]}) \in \Theta^l \text{ for some } \boldsymbol{\theta}_{[\sim(j+1)]} \right\} \quad .$$

Also, let $\mathbf{1}(y \in \Omega)$ denote the indicator function that equals 1 if $y \in \Omega$ and 0 otherwise.

Specifically, one follows an algorithm of Berger and Bernardo (1992b). Towards this, first define $|\mathbf{h}_j(\boldsymbol{\theta})|$ following Datta and Ghosh (1996) as

$$|\mathbf{h}_j(\boldsymbol{\theta})| = \frac{|\mathbf{I}_{[\sim j-1, j-1]}|}{|\mathbf{I}_{[\sim jj]}|}, \quad j = 1, \dots, m$$

where

$$\mathbf{I}_{[\sim jj]} = ((\mathbf{I}_{ik}))_{i=j+1, \dots, m, k=j+1, \dots, m} \quad .$$

The algorithm is now described as follows :

Start. Define

$$\begin{aligned} \pi_m^l(\boldsymbol{\theta}_{[\sim(m-1)]}|\boldsymbol{\theta}_{[m-1]}) &= \pi_m^l(\boldsymbol{\theta}_{(m)}|\boldsymbol{\theta}_{[m-1]}) \\ &= \frac{|\mathbf{h}_m(\boldsymbol{\theta})|^{1/2} \mathbf{1}(\boldsymbol{\theta}_{(m)} \in \Theta^l(\boldsymbol{\theta}_{[m-1]}))}{\int_{\boldsymbol{\theta}_{(m)} \in \Theta^l(\boldsymbol{\theta}_{[m-1]})} |\mathbf{h}_m(\boldsymbol{\theta})|^{1/2} d\boldsymbol{\theta}_{(m)}} \end{aligned} \quad (1.1)$$

Iteration. For $j = m-1, m-2, \dots, 1$, define

$$\begin{aligned} &\pi_j^l(\boldsymbol{\theta}_{[\sim(j-1)]}|\boldsymbol{\theta}_{[j-1]}) \\ &= \frac{\pi_{j+1}^l(\boldsymbol{\theta}_{[\sim j]}|\boldsymbol{\theta}_{[j]}) \exp \left\{ \frac{1}{2} E_j^l \left[(\log |\mathbf{h}_j(\boldsymbol{\theta})|) |\boldsymbol{\theta}_{[j]} \right] \right\} \mathbf{1}(\boldsymbol{\theta}_{(j)} \in \Theta^l(\boldsymbol{\theta}_{[j-1]}))}{\int_{\boldsymbol{\theta}_{(j)} \in \Theta^l(\boldsymbol{\theta}_{[j-1]})} \exp \left\{ \frac{1}{2} E_j^l \left[(\log |\mathbf{h}_j(\boldsymbol{\theta})|) |\boldsymbol{\theta}_{[j]} \right] \right\} d\boldsymbol{\theta}_{(j)}} \end{aligned} \quad (1.2)$$

where

$$E_j^l \left(g(\boldsymbol{\theta}) | \boldsymbol{\theta}_{[j]} \right) = \int g(\boldsymbol{\theta}) \pi_{j+1}^l \left(\boldsymbol{\theta}_{[\sim j]} | \boldsymbol{\theta}_{[j]} \right) d\boldsymbol{\theta}_{[\sim j]} \quad . \quad (1.3)$$

Note that it is easy to check, by integrating in succession over $\boldsymbol{\theta}_{(m)}, \boldsymbol{\theta}_{(m-1)}, \dots, \boldsymbol{\theta}_{(j)}$, that π_j^l defines a probability distribution. For $j = 1$, interpret $\boldsymbol{\theta}_{[\sim 0]}$ as $\boldsymbol{\theta}$, $\boldsymbol{\theta}_{[0]}$ as vacuous, and write

$$\pi^l(\boldsymbol{\theta}) = \pi_1^l \left(\boldsymbol{\theta}_{[\sim 0]} | \boldsymbol{\theta}_{[0]} \right) . \quad (1.4)$$

Finish. Define the m -group reference prior by

$$\pi(\boldsymbol{\theta}) = \lim_{l \rightarrow \infty} \frac{\pi^l(\boldsymbol{\theta})}{\pi^l(\boldsymbol{\theta}^*)} \quad , \quad (1.5)$$

where $\boldsymbol{\theta}^*$ is some point in Θ^l .

Note that if the integrals and expectation in (1.1) and (1.2) are finite when the ' l ' is removed, that is, when Θ^l is replaced by Θ everywhere, then the reference prior is defined simply by π_1 , so that (1.5) is not needed.

In particular, suppose that $|\mathbf{h}_j(\boldsymbol{\theta})|$ depends only on $\boldsymbol{\theta}_{[j]}$, $j = 1, \dots, m$. Then, the m -group reference prior is given by

$$\pi^l(\boldsymbol{\theta}) = \left(\prod_{i=1}^m \frac{|\mathbf{h}_i(\boldsymbol{\theta})|^{1/2}}{\int_{\boldsymbol{\theta}_{(i)} \in \Theta^l(\boldsymbol{\theta}_{[i-1]})} |\mathbf{h}_i(\boldsymbol{\theta})|^{1/2} d\boldsymbol{\theta}_{(i)}} \right) \mathbf{1} \left(\boldsymbol{\theta} \in \Theta^l \right) .$$

Also Datta and M.Ghosh (1995) simplified the calculation of the reference priors under some conditions. Assume that the Fisher information matrix $\mathbf{I}(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$ is

$$\mathbf{I}(\boldsymbol{\theta}) = \text{block diagonal}(\mathbf{h}_1(\boldsymbol{\theta}), \dots, \mathbf{h}_m(\boldsymbol{\theta})),$$

where $\mathbf{h}_j(\boldsymbol{\theta})$ is $n_j \times n_j$ (may not be diagonal). Let $\boldsymbol{\theta}_{(j)}^C = (\boldsymbol{\theta}_{(1)}, \dots, \boldsymbol{\theta}_{(j-1)}, \boldsymbol{\theta}_{(j+1)}, \dots, \boldsymbol{\theta}_{(m)})$, $j = 1, \dots, m$. Assume that

$$|\mathbf{h}_j(\boldsymbol{\theta})| = h_{j1}(\boldsymbol{\theta}_{(j)}) h_{j2}(\boldsymbol{\theta}_{(j)}^C)$$

for nonnegative functions h_{j1} and h_{j2} . Then,

$$\pi(\boldsymbol{\theta}) = \prod_{j=1}^m h_{j1}^{1/2}(\boldsymbol{\theta}_{(j)}).$$

A somewhat different criterion for developing noninformative priors is based on matching the posterior coverage probability of Bayesian credible intervals for a real-valued parameter of interest with the corresponding frequentist coverage probability. Such matching is accomplished through either (a) posterior quantiles, (b) the highest posterior density (HPD) regions, or (c) the inversion of likelihood ratio and related statistics. In many literatures, (a), (b) and (c) have received attention in decreasing order of importance.

We shall now discuss these matching concepts in greater details. We begin with the discussion of (a). Let $\{\mathbf{X}_i, i \geq 1\}$ be a sequence of independent and identically distributed possibly vector-valued random variables with common density $f(\mathbf{x}; \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$ belongs to some open subset of $\mathcal{R}^p$, and θ_1 is the parameter of interest. Let $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)^T$, where n is the sample size, and denote by $P^\pi\{\cdot|\mathbf{X}\}$ the posterior probability measure for $\boldsymbol{\theta}$ under a prior $\pi(\boldsymbol{\theta})$. Suppose $\theta_1^{(1-\alpha)}(\pi, \mathbf{X})$ denotes an approximate $(1 - \alpha)$ th posterior quantile which satisfies $P^\pi\{\theta_1 \leq \theta_1^{(1-\alpha)}(\pi, \mathbf{X})|\mathbf{X}\} = 1 - \alpha + o(n^{-\frac{1}{2}})$. We seek to characterize the class of priors $\pi(\boldsymbol{\theta})$ such that $P\{\theta_1 \leq \theta_1^{(1-\alpha)}(\pi, \mathbf{X})|\boldsymbol{\theta}\} = 1 - \alpha + o(n^{-\frac{1}{2}})$. Priors π satisfying this property will be called first order probability matching priors.

A brief introduction to matching priors is given in Lindley (1958), but their systematic study began with Welch and Peers (1963) who considered development of such priors in the absence of nuisance parameters, for $p = 1$. These authors showed in this case that the unique prior satisfying the first order matching property is Jeffreys' prior. Peers (1965) followed up this study by including nuisance parameters. Writing $\mathbf{I}[\equiv \mathbf{I}(\boldsymbol{\theta})] = (I_{ij})$ as the Fisher information matrix per unit observation

and $I^{-1}[\equiv I^{-1}(\boldsymbol{\theta})] = ((I^{ij}))$, he characterized the class of priors π satisfying the first order matching property as solutions of the partial differential equation

$$\Delta_1(\pi, \boldsymbol{\theta}) = \sum_{j=1}^p \frac{\partial}{\partial \theta_j} \{ (I^{11})^{-\frac{1}{2}} I^{j1} \pi(\boldsymbol{\theta}) \} = 0 \quad . \quad (1.6)$$

We also refer to Stein (1985) in this context. For $p \geq 2$, one may be interested in priors that are first order probability matching for each component of $\boldsymbol{\theta}$. Starting from equations like (1.6) separately for each $\theta_1, \dots, \theta_p$, Peers (1965) gave necessary and sufficient conditions on the model for the existence of such priors.

As it turns out, Jeffreys' prior $\pi^J(\boldsymbol{\theta}) \propto |\mathbf{I}(\boldsymbol{\theta})|^{\frac{1}{2}}$ need not always satisfy (1.6) even though other priors satisfying (1.6) could exist. An important simplification of (1.6) is due to Tibshirani (1989) in the case when θ_1 is orthogonal to the nuisance parameters $(\theta_2, \dots, \theta_p)$, that is, when $I^{j1} = 0$, identically in $\boldsymbol{\theta}$ ($2 \leq j \leq p$). In this case, the class of first order probability matching priors $\pi(\boldsymbol{\theta})$ is characterized by

$$\pi(\boldsymbol{\theta}) = I_{11}^{1/2} g(\theta_2, \dots, \theta_p), \quad (1.7)$$

where g is an arbitrary positive function.

The idea of matching was pursued further in later papers by Nicolau (1993) and Datta and M.Ghosh (1995). As it is easily seen, there are infinitely many first order probability matching priors in the presence of nuisance parameters. With the intent of narrowing down the selection of priors, Mukerjee and Dey (1993) considered priors with Bayesian credible intervals and frequentist confidence intervals matching up to $o(n^{-1})$, instead of $o(n^{-\frac{1}{2}})$, for the case of single nuisance parameter. These results were later generalized by Mukerjee and Ghosh (1997) in the presence of several nuisance parameters. The findings of these authors indicate that second order probability matching priors need not always exist, while sometimes all first order probability matching priors also meet the second order matching criterion. Barring these extreme

situations, second order matching priors help to narrow down the selection of priors and indeed often lead to a unique prior within the class of first order probability matching priors. In order to develop such priors, let $\tau^{jr} = I^{j1}I^{r1}/I^{11}$ and $L_{jrs} = E_{\theta}\{\partial^3 \log f(\mathbf{X}_1; \theta)/\partial\theta_j\partial\theta_r\partial\theta_s\}$, $1 \leq j, r, s \leq p$. Then Mukerjee and Ghosh (1997) showed that a prior π is second order probability matching based on the posterior quantiles if and only if it satisfies (1.6) and, in addition, satisfies a second partial differential equation $\Delta_2(\pi, \theta) = 0$, where

$$\begin{aligned} \Delta_2(\pi, \theta) = & \sum_{j=1}^p \sum_{r=1}^p \frac{\partial^2}{\partial\theta_j\partial\theta_r} \{\tau^{jr}\pi(\theta)\} \\ & - \frac{1}{3} \sum_{v=1}^p \sum_{j=1}^p \sum_{r=1}^p \sum_{s=1}^p \frac{\partial}{\partial\theta_v} \{L_{jrs}\tau^{jr}(3I^{sv} - 2\tau^{sv})\pi(\theta)\} . \end{aligned} \quad (1.8)$$

Under parametric orthogonality, (1.8) simplifies to

$$\frac{1}{6}g(\theta_2, \dots, \theta_p) \frac{\partial}{\partial\theta_1} (I_{11}^{-3/2} L_{1,1,1}) + \sum_{v=2}^p \sum_{s=2}^p \frac{\partial}{\partial\theta_v} \{I_{11}^{-1/2} L_{11s} I^{sv} g(\theta_2, \dots, \theta_p)\} = 0 , \quad (1.9)$$

where $L_{1,1,1} = E_{\theta}[\{\partial \log f(\mathbf{X}_1; \theta)/\partial\theta_1\}^3]$. With only one nuisance parameter, (1.9) reduces to

$$\frac{1}{6}g(\theta_2) \frac{\partial}{\partial\theta_1} (I_{11}^{-3/2} L_{1,1,1}) + \frac{\partial}{\partial\theta_2} \{I_{11}^{-1/2} L_{112} I_{22}^{-1} g(\theta_2)\} = 0, \quad (1.10)$$

which is due to Mukerjee and Dey (1993). For $p = 1$, that is when there are no nuisance parameters, it follows from (1.9) that Jeffreys' prior is second order probability matching if and only if $I_{11}^{-3/2} L_{1,1,1}$ is free from θ_1 .

Probability matching priors are invariant under parametric transformation (cf. Datta and Ghosh, 1996 ; Mukerjee and Ghosh, 1997), while reference priors are invariant only under parametric transformation within each group.

Probability matching priors for specific problems are investigated by a few authors. Ghosh, Carlin and Srivastava (1995) derived and studied a class of first order

probability matching priors and a complete catalog of reference priors of the univariate linear calibration problem. Ghosh and Yang (1996) derived a class of first order probability matching priors for two sample normal problem. Sun and Ye (1996) considered a two-parameter exponential family with first and second order probability matching priors. Garvan and Ghosh (1997) studied probability matching and reference priors in the more general case of dispersion models of Jørgensen (1992).

Next, we discuss the criteria (b) and (c).

DiCiccio and Stern (1994) and Ghosh and Mukerjee (1995) discussed matching through the HPD region. Specifically, if $\tilde{\pi}$ denotes the posterior distribution of θ_1 under a prior π , and $k_\alpha \equiv k_\alpha(\pi, \mathbf{X})$ is such that

$$P^\pi \left[\tilde{\pi}(\theta_1 | \mathbf{X}) \geq k_\alpha | \mathbf{X} \right] = 1 - \alpha + o(n^{-1}) \quad , \quad (1.11)$$

then the HPD region for θ_1 with posterior coverage probability $1 - \alpha + o(n^{-1})$ is given by

$$H_\alpha(\pi, \mathbf{X}) = \{\theta_1 : \tilde{\pi}(\theta_1, \dots, \theta_p | \mathbf{X}) \geq k_\alpha\} \quad . \quad (1.12)$$

DiCiccio and Stern (1994) and Ghosh and Mukerjee (1995) characterized priors π for which

$$P \left[\theta_1 \in H_\alpha(\pi, \mathbf{X}) | \theta_1, \dots, \theta_p \right] = 1 - \alpha + o(n^{-1}) \quad , \quad (1.13)$$

for all $\theta_1, \dots, \theta_p$ and all $\alpha \in (0, 1)$. They found necessary and sufficient conditions under which π satisfies (1.13).

Yet a third criterion of matching is via inversion of posterior Bartlett corrected likelihood ratio statistics. This is investigated in Ghosh and Mukerjee (1991) and in DiCiccio and Stern (1994). Specifically, these authors have found priors under which the Bayesian and frequentist Bartlett corrections for the likelihood ratio test statistic

differ by $o(1)$ when the sample size tends to infinity. Similar results are found in Ghosh and Mukerjee (1992) for conditional likelihood ratio test statistics.

As mentioned in Ghosh (1994, p 86), there are usually four criteria associated with the development of noninformative priors. These are (i) maximization of entropy or minimization of information; (ii) matching asymptotically the coverage probabilities of Bayesian credible sets with the corresponding frequentist probabilities; (iii) principle of group invariance; and (iv) minimaxity of Bayesian procedures. Of these, (i) and (ii) have found the widest applicability in the Bayesian literature. The reference priors are very satisfactory under criterion (i), while probability matching priors, as the name implies, are ideal for (ii).

The utility of noninformative priors has always been questioned by subjectivists. Yet one cannot deny their pragmatic appeal. Indeed, the wider acceptance of Bayesian techniques in recent years both in the theory and in the practice of statistics can be partly attributed to the fact that even with little or vague prior information, Bayesian techniques can often be used successfully by employing noninformative priors. Reid (1995) and Kass and Wasserman (1996) contain excellent reviews of noninformative priors.

1.2 The Subject of This Dissertation

In Chapter 2, we consider the Behrens-Fisher problem which involves inference about the difference of two normal means when the ratio of the two variances is unknown. The class of second order probability matching priors, the reference priors, and the Jeffreys' general rule prior are derived and are compared with the Jeffreys' independent prior via simulation study. One particular second order probability matching prior is recommended for use. It turns out that this second order probability matching prior satisfies the matching criteria (b) and (c). Our simulation study

indicates that second order probability matching priors perform better than the other priors in terms of target coverage probabilities in a frequentist sense even for small and moderate sample sizes.

Chapter 3 considers the interval estimation of the common mean of several normal populations. The class of first order probability matching priors, the reference priors, Jeffreys' prior, and a second order probability matching prior are derived. It turns out that the second order probability matching prior also satisfies the matching criteria (b) and (c). Equal tailed and HPD credible intervals are compared against several frequentist confidence intervals.

Chapter 4 is devoted to Bayesian inference involving ratios of regression coefficients. A nontrivial orthogonal reparameterization is introduced which facilitates the development of default Bayesian priors such as the reference priors and probability matching priors in the general case of multiple regression models. As special cases, we have considered the Fieller-Creasy problem, linear calibration problem, parallel-line bioassay and slope-ratio bioassay.

Finally, in Chapter 5, we summarize the result of this dissertation and propose several topics for future research.

CHAPTER 2 THE BEHRENS-FISHER PROBLEM

2.1 Introduction

The Behrens-Fisher problem is that of testing whether the means of two normal populations are equal without necessarily assuming equality of the variances. An essentially equivalent problem is that of finding a confidence interval for the difference of two normal means.

Specifically, suppose that two independent random samples of sizes $n_1 (\geq 2)$ and $n_2 (\geq 2)$ are drawn from two normal populations with respective means μ_1 and μ_2 , and respective variances σ_1^2 and σ_2^2 . The corresponding sample means are denoted by $\bar{X}_1$ and $\bar{X}_2$, and sample variances with denominators $n_1 - 1$ and $n_2 - 1$ are denoted by S_1^2 and S_2^2 .

If the variance ratio $\sigma_1^2/\sigma_2^2 = \eta$ is assumed to be known (and, in particular, when $\eta = 1$), then the pivot

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{[(n_1 - 1)S_1^2\eta^{-1} + (n_2 - 1)S_2^2]^{\frac{1}{2}}} \left(\frac{n_1 + n_2 - 2}{\eta m_1^{-1} + n_2^{-1}} \right)^{\frac{1}{2}} \quad (2.1)$$

has a t -distribution with $n_1 + n_2 - 2$ degrees of freedom yielding the $100(1 - \alpha)\%$ confidence interval

$$\bar{X}_1 - \bar{X}_2 \pm t_{\alpha/2; n_1 + n_2 - 2} \left[\frac{(n_1 - 1)S_1^2\eta^{-1} + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \right]^{\frac{1}{2}} (\eta m_1^{-1} + n_2^{-1})^{\frac{1}{2}} \quad (2.2)$$

for $\mu_1 - \mu_2$. Throughout, we will use the symbol $t_{\alpha; \nu}$ for the upper $100(1 - \alpha)\%$ point of Student's t with ν degrees of freedom. The Behrens-Fisher problem arises when the variance ratio σ_1^2/σ_2^2 is unknown.

The above inference problem has received considerable attention for several decades. Here the fiducial interval of Fisher differs drastically from the Neyman-Pearson confidence interval. This issue is discussed extensively in Kendall and Stuart (1967). Scheffé (1970) and Lee and Gurland (1975) provide reviews of various solutions to the problem.

Behrens (1929) was the first to offer a solution to the Behrens-Fisher problem in a testing context, namely $H_0 : \mu_1 = \mu_2$ against $H_1 : \mu_1 \neq \mu_2$. Later, Fisher (1935) pointed out that this solution could be justified using the fiducial theory of inference. From a frequentist perspective, Bartlett (1936) noted that Behrens' test has a size different from what was originally intended. Inverting this test into an interval, this amounts to the fact that the coverage probability of the confidence interval for $\mu_1 - \mu_2$ is different from the specified confidence coefficient. Behrens' interval is based on the pivot

$$D = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{(n_1^{-1}S_1^2 + n_2^{-1}S_2^2)^{1/2}} \quad .$$

In contrast, Welch (1947) provided a test having type I error very close to the nominal value throughout the parameter space. Thus, the resulting confidence interval for $\mu_1 - \mu_2$ has coverage probability very nearly equal to the target confidence coefficient. Welch's procedure is also based on the pivot D , but he uses different significance points. These calculations were later extended by Aspin (1948).

Fisher (1956) criticized Welch's test in a way that amounts to showing the existence of negatively biased relevant subset in the sense of Buehler (1959). For instance, as shown in Robinson (1982), $P[|D| > a | S_1/S_2 = 1] \geq P[|t_{n_1+n_2-2}| > a]$ for all $a > 0$. Thus, the set where $S_1/S_2 = 1$ is a relevant subset, confidence intervals based on the Welch-Aspin test covers the true value of $\mu_1 - \mu_2$ less often than what the nominal value suggests. On the other hand, as shown by Robinson (1976), negatively biased relevant subsets do not exist for Behrens' solution.

Jeffreys (1961) pointed out that a Bayesian calculation based on the so-called independent prior $\pi^{II}(\mu_1, \mu_2, \sigma_1, \sigma_2) \propto (\sigma_1 \sigma_2)^{-1}$ yields a credible interval for $\mu_1 - \mu_2$ which is algebraically equivalent to the fiducial interval of Fisher, although the interpretations are necessarily different. Throughout, we will refer to this prior as Jeffreys' *independent* prior. The same prior is mentioned in Cox and Hinkley (1974). It is worth pointing out here that the above Jeffreys' independent prior is different from Jeffreys' general rule prior which is the positive square root of the determinant of the Fisher information matrix, and in this example, is proportional to $(\sigma_1 \sigma_2)^{-2}$.

Thus, while the prior $\pi^{II}(\mu_1, \mu_2, \sigma_1, \sigma_2) \propto (\sigma_1 \sigma_2)^{-1}$ lends support to the fiducial solution (though does not support the fiducial reasoning) both from Bayesian and conditional frequentist perspectives, it is not clear whether this prior can necessarily be justified from the usual unconditional frequentist criterion. In Section 2 of this chapter, we provide some justification of this prior from a frequentist angle. More important, we find a new prior in this section, credible intervals based on which match asymptotically the corresponding frequentist coverage probabilities more accurately than Jeffreys' independent prior. Though this matching can be justified only asymptotically, our simulation results indicate that this is indeed achieved for small or moderate sample sizes as well. This prior is also satisfactory from a conditional frequentist perspective.

In Section 2 of this chapter, we develop first and second order probability matching priors for the Behrens-Fisher problem when the parameter of interest is $\mu_1 - \mu_2$. It is shown that the Jeffreys' independent prior is a first order probability matching prior, but is not a second order probability matching prior. An alternate prior which is second order probability matching prior is derived in this section. This new prior is justified also by some alternate matching criteria such as HPD matching, and matching via inversion of likelihood ratio test statistics. Also, in this section, we bring in some alternate criteria for the development of noninformative priors. In

particular, we find the “two-group” and “one-at-a-time” reference priors as developed in Bernardo (1979) and Berger and Bernardo (1989,1992a,b). It turns out that the prior proposed by Jeffreys is one-at-a-time reference prior based on some particular ordering of the parameters.

In Section 3, the propriety of posteriors under the different priors developed in Section 2 is established. Certain other properties of these posteriors such as symmetry and unimodality are also proved in this section. Also, a small simulation study is performed which shows that the second order probability matching prior developed in this chapter matches the frequentist target coverage probabilities better than Jeffreys’ independent prior especially for small or moderate samples. Section 4 provides justification of the former from a conditional frequentist perspective. In particular, it is pointed out that unlike the Welch-Aspin solution, the credible intervals under the proposed second order matching prior, though satisfactory from a frequentist perspective, do not admit negatively biased relevant subsets. Some concluding remarks are made in Section 5.

In comparison with Jeffreys’ independent prior or the one-at-a-time reference prior, the new prior proposed in this chapter seems to be more appropriate under the HPD criterion, especially for achieving the Bayes-frequentist synthesis for small and moderate sample sizes. This is evidenced in the numerical findings of Section 4. Clearly, there are situations such as the Fieller-Creasy (the ratio of normal means) problem where the Bayes-frequentist synthesis may not be appropriate. There the frequentist confidence set may be two disjoint unbounded sets, or sometimes even the whole real line. However, such difficulty is not encountered in the Behrens-Fisher problem by frequentist procedures, and as such, the proposed Bayesian method does not suffer any frequentist shortcomings. Our findings are also quite in conformity with Berger and Bernardo (1989) who recommended evaluation of noninformative priors by good frequentist properties. Finally, Jeffreys’ independent prior and the

newly developed prior are both quite satisfactory from a conditional frequentist perspective.

2.2 Development of Noninformative Priors

2.2.1 Probability Matching Priors

Let $X_{11}, \dots, X_{1n_1}, X_{21}, \dots, X_{2n_2}$ be mutually independent with X_{ij} ($j = 1, \dots, n_i; i = 1, 2$) *iid* $\mathcal{N}(\mu_i, \sigma_i^2)$. We write $\mathbf{X}_1 = (X_{11}, \dots, X_{1n_1})$, $\mathbf{X}_2 = (X_{21}, \dots, X_{2n_2})$, $n_i = N\rho_i$ ($i = 1, 2$), where $\rho_1 + \rho_2 = 1$. The parameter of interest is $\theta_1 = \mu_1 - \mu_2$. For any given prior π , we denote the posterior by $P^\pi(\cdot | \mathbf{X}_1, \mathbf{X}_2)$. The first objective is to find the class of first order probability matching priors when θ_1 is the parameter of interest. In this example, this amounts to finding a π such that

$$P^\pi \left[\theta_1 \leq \theta_1^{(1-\alpha)}(\pi, \mathbf{X}_1, \mathbf{X}_2) \middle| \mu_1, \mu_2, \sigma_1, \sigma_2 \right] = 1 - \alpha + o(N^{-u}) \quad (2.3)$$

as $N \rightarrow \infty$ for some $u > 0$, where $\theta_1^{(1-\alpha)}(\pi, \mathbf{X}_1, \mathbf{X}_2)$ denotes the $(1 - \alpha)$ th posterior quantile of θ_1 based on the prior π .

We begin by noting that the Fisher information matrix $\mathbf{I}(\mu_1, \mu_2, \sigma_1, \sigma_2)$ is given by

$$\mathbf{I}(\mu_1, \mu_2, \sigma_1, \sigma_2) = \text{Diagonal} \left(\frac{n_1}{\sigma_1^2}, \frac{n_2}{\sigma_2^2}, \frac{2n_1}{\sigma_1^2}, \frac{2n_2}{\sigma_2^2} \right) .$$

Next we propose the one-to-one reparameterization

$$\begin{aligned} \theta_1 &= \mu_1 - \mu_2 \\ \theta_2 &= \frac{n_1 \sigma_1^{-2} \mu_1 + n_2 \sigma_2^{-2} \mu_2}{n_1 \sigma_1^{-2} + n_2 \sigma_2^{-2}} \\ \theta_3 &= \sigma_1 \\ \theta_4 &= \sigma_2 \end{aligned}$$

which transforms the Fisher information matrix to

$$\mathbf{I}(\boldsymbol{\theta}) = \begin{bmatrix} (\theta_3^2/n_1 + \theta_4^2/n_2)^{-1} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{I}_{22} \end{bmatrix} \quad (2.4)$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)$ and

$$\mathbf{I}_{22} = \begin{bmatrix} \frac{n_1}{\theta_3^2} + \frac{n_2}{\theta_4^2} & \frac{2\theta_1}{\theta_3(\theta_3^2/n_1 + \theta_4^2/n_2)} & -\frac{2\theta_1}{\theta_4(\theta_3^2/n_1 + \theta_4^2/n_2)} \\ \frac{2\theta_1}{\theta_3(\theta_3^2/n_1 + \theta_4^2/n_2)} & \frac{2n_1}{\theta_3^2} + \frac{4\theta_1^2\theta_4^2}{n_1n_2(\theta_3^2/n_1 + \theta_4^2/n_2)^3} & -\frac{4\theta_1^2\theta_3\theta_4}{n_1n_2(\theta_3^2/n_1 + \theta_4^2/n_2)^3} \\ -\frac{2\theta_1}{\theta_4(\theta_3^2/n_1 + \theta_4^2/n_2)} & -\frac{4\theta_1^2\theta_3\theta_4}{n_1n_2(\theta_3^2/n_1 + \theta_4^2/n_2)^3} & \frac{2n_2}{\theta_4^2} + \frac{4\theta_1^2\theta_3^2}{n_1n_2(\theta_3^2/n_1 + \theta_4^2/n_2)^3} \end{bmatrix}. \quad (2.5)$$

In order to derive (2.5), we may note that the transformation is equivalent to

$$\begin{aligned} \mu_1 &= \theta_2 + \frac{n_2\theta_1\theta_3^2}{n_2\theta_3^2 + n_1\theta_4^2} \\ \mu_2 &= \theta_2 - \frac{n_1\theta_1\theta_4^2}{n_2\theta_3^2 + n_1\theta_4^2} \\ \sigma_1 &= \theta_3 \\ \sigma_2 &= \theta_4 \end{aligned}$$

so that the Jacobian matrix is given by

$$\mathbf{J} = \begin{bmatrix} \frac{n_2\theta_3^2}{n_2\theta_3^2 + n_1\theta_4^2} & -\frac{n_1\theta_4^2}{n_2\theta_3^2 + n_1\theta_4^2} & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \frac{2n_1n_2\theta_1\theta_3\theta_4^2}{(n_2\theta_3^2 + n_1\theta_4^2)^2} & \frac{2n_1n_2\theta_1\theta_3\theta_3^2}{(n_2\theta_3^2 + n_1\theta_4^2)^2} & 1 & 0 \\ -\frac{2n_1n_2\theta_1\theta_3^2\theta_4}{(n_2\theta_3^2 + n_1\theta_4^2)^2} & -\frac{2n_1n_2\theta_1\theta_4^2\theta_3}{(n_2\theta_3^2 + n_1\theta_4^2)^2} & 0 & 1 \end{bmatrix}.$$

Now $\mathbf{I}(\boldsymbol{\theta})$ is found as $\mathbf{I}(\boldsymbol{\theta}) = \mathbf{J}\mathbf{I}(\boldsymbol{\mu})\mathbf{J}^T$, where $\mathbf{I}(\boldsymbol{\mu})$ expressed in terms of $\boldsymbol{\theta}$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \sigma_1, \sigma_2)$. By (2.4), θ_1 is orthogonal to $(\theta_2, \theta_3, \theta_4)$ in the sense of Cox and Reid (1987). Now, from Tibshirani (1989) and Datta and J.K.Ghosh (1995), the class of first order probability matching priors is characterized by

$$\pi^F(\boldsymbol{\theta}) \propto (\theta_3^2/n_1 + \theta_4^2/n_2)^{-\frac{1}{2}} g(\theta_2, \theta_3, \theta_4), \quad (2.6)$$

where g is any arbitrary function differentiable in its arguments.

Due to the arbitrariness of g , the class of priors given in (2.6) contains infinitely many members. In order to narrow down the selection of priors within this class, we consider *second order probability matching priors* (see Mukerjee and Dey, 1993; Mukerjee and Ghosh, 1997).

To this end, first rewrite the likelihood as

$$\begin{aligned} L(\boldsymbol{\theta}) &\propto \theta_3^{-n_1} \theta_4^{-n_2} \exp \left[-\frac{1}{2\theta_3^2} \sum_{i=1}^{n_1} \left(x_{1i} - \theta_2 - \frac{\theta_1 \theta_3^2 / n_1}{\theta_3^2 / n_1 + \theta_4^2 / n_2} \right)^2 \right] \\ &\times \exp \left[-\frac{1}{2\theta_4^2} \sum_{j=1}^{n_2} \left(x_{2j} - \theta_2 + \frac{\theta_1 \theta_4^2 / n_2}{\theta_3^2 / n_1 + \theta_4^2 / n_2} \right)^2 \right]. \end{aligned} \quad (2.7)$$

Writing $\mathbf{I}(\boldsymbol{\theta}) = ((I_{ij}))$, $i, j = 1, \dots, 4$, and following Mukerjee and Ghosh (1997), a prior given in (2.6) is a second order probability matching prior if and only if g satisfies

$$\frac{1}{6} g(\theta_2, \theta_3, \theta_4) \frac{\partial}{\partial \theta_1} (I_{11}^{-\frac{3}{2}} L_{1,1,1}) + \sum_{\nu=2}^4 \sum_{s=2}^4 \frac{\partial}{\partial \theta_\nu} \{ I_{11}^{-\frac{1}{2}} L_{11s} I^{s\nu} g(\theta_2, \theta_3, \theta_4) \} = 0 \quad (2.8)$$

where $L_{1,1,1} = E_{\boldsymbol{\theta}} \left(\frac{\partial \log L}{\partial \theta_1} \right)^3$ and $L_{11s} = E_{\boldsymbol{\theta}} \left(\frac{\partial^3 \log L}{\partial \theta_1^2 \partial \theta_s} \right)$, $s = 2, 3, 4$. From (2.7), after considerable algebra, we have

$$\begin{aligned} L_{1,1,1} &= 0 \\ L_{112} &= 0 \\ L_{113} &= \frac{2\theta_3}{n_1} \left(\frac{\theta_3^2}{n_1} + \frac{\theta_4^2}{n_2} \right)^{-2} \\ L_{114} &= \frac{2\theta_4}{n_2} \left(\frac{\theta_3^2}{n_1} + \frac{\theta_4^2}{n_2} \right)^{-2}. \end{aligned}$$

Finally, writing the inverse matrix of $\mathbf{I}(\boldsymbol{\theta})$ as $\mathbf{I}^{-1}(\boldsymbol{\theta}) = ((I^{ij}))$, $i, j = 1, 2, 3, 4$, the resulting inverse matrix $\mathbf{I}^{-1}(\boldsymbol{\theta})$ is given by

$$\mathbf{I}^{-1}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\theta_3^2}{n_1} + \frac{\theta_4^2}{n_2} & 0 & 0 & 0 \\ 0 & I^{22} & -\frac{n_2 \theta_1 \theta_3^3 \theta_4^2}{(n_2 \theta_3^2 + n_1 \theta_4^2)^2} & \frac{n_1 \theta_1 \theta_3^2 \theta_4^3}{(n_2 \theta_3^2 + n_1 \theta_4^2)^2} \\ 0 & -\frac{n_2 \theta_1 \theta_3^3 \theta_4^2}{(n_2 \theta_3^2 + n_1 \theta_4^2)^2} & \frac{\theta_3^2}{2n_1} & 0 \\ 0 & \frac{n_1 \theta_1 \theta_3^2 \theta_4^3}{(n_2 \theta_3^2 + n_1 \theta_4^2)^2} & 0 & \frac{\theta_4^2}{2n_2} \end{bmatrix},$$

where $I^{22} = \frac{\theta_3^2 \theta_4^2}{n_2 \theta_3^2 + n_1 \theta_4^2} + \frac{2n_1 n_2 (n_1 + n_2) \theta_1^2 \theta_3^4 \theta_4^4}{(n_2 \theta_3^2 + n_1 \theta_4^2)^4}$.

With these substitutions (2.8) reduces to

$$n_1^{-2} \frac{\partial}{\partial \theta_3} \left[\left(\frac{\theta_3^2}{n_1} + \frac{\theta_4^2}{n_2} \right)^{-\frac{3}{2}} \theta_3^3 g(\theta_2, \theta_3, \theta_4) \right] + n_2^{-2} \frac{\partial}{\partial \theta_4} \left[\left(\frac{\theta_3^2}{n_1} + \frac{\theta_4^2}{n_2} \right)^{-\frac{3}{2}} \theta_4^3 g(\theta_2, \theta_3, \theta_4) \right] = 0 \quad (2.9)$$

Consider the class of functions $g(\cdot)$ such that

$$g(\theta_2, \theta_3, \theta_4) = k(\theta_2) h(\theta_3, \theta_4) \left(\frac{\theta_3^2}{n_1} + \frac{\theta_4^2}{n_2} \right)^{\frac{3}{2}} \theta_3^{-3} \theta_4^{-3}.$$

Then, (2.9) reduces to

$$n_1^{-2} \theta_4^{-3} \frac{\partial}{\partial \theta_3} h + n_2^{-2} \theta_3^{-3} \frac{\partial}{\partial \theta_4} h = 0$$

Equivalently,

$$\frac{\theta_3^3}{n_1^2} \frac{\partial}{\partial \theta_3} h + \frac{\theta_4^3}{n_2^2} \frac{\partial}{\partial \theta_4} h = 0 \quad (2.10)$$

A general class of solutions to (2.10) is given by

$$h(\theta_3, \theta_4) = h^* \left(\frac{n_1^2}{\theta_3^2} - \frac{n_2^2}{\theta_4^2} \right),$$

where $h^*(\cdot)$ is an arbitrary function differentiable in its arguments. Hence, a general class of solutions to (2.9) is given by

$$g(\theta_2, \theta_3, \theta_4) = \theta_3^{-3} \theta_4^{-3} \left(\frac{\theta_3^2}{n_1} + \frac{\theta_4^2}{n_2} \right)^{\frac{3}{2}} u \left(\frac{n_1^2}{\theta_3^2} - \frac{n_2^2}{\theta_4^2} \right),$$

where $u(\cdot)$ is an arbitrary differentiable function. However, not every solution is permissible in the construction of priors. For instance, since the prior has to be nonnegative, any function of the form $(n_1^2/\theta_3^2 - n_2^2/\theta_4^2)^{2k+1}$, k being a positive integer, needs to be excluded from consideration. However, functions of the form $(n_1^2/\theta_3^2 - n_2^2/\theta_4^2)^{2k}$, k being a nonnegative integer, indeed lead to proper posteriors. But these posteriors become more and more complex as k gets larger. Also, there does not seem

to be any improvement in the coverage probabilities with these complex posteriors. Thus, we have chosen $u(\cdot)$ to be a constant function. The resulting second order probability matching prior is then given by

$$\pi^S(\boldsymbol{\theta}) \propto \theta_3^{-3} \theta_4^{-3} (\theta_3^2/n_1 + \theta_4^2/n_2) \quad . \quad (2.11)$$

In the original parameterization this prior becomes

$$\pi^S(\mu_1, \mu_2, \sigma_1, \sigma_2) \propto \sigma_1^{-3} \sigma_2^{-3} (\sigma_1^2/n_1 + \sigma_2^2/n_2) \quad . \quad (2.12)$$

Remark Due to the invariance of matching priors (Datta and Ghosh, 1996; Mukerjee and Ghosh, 1997), $\pi^S(\boldsymbol{\theta})$ remains a second order probability matching prior for θ_1 under any one-to-one reparameterization of $\boldsymbol{\theta}$ leaving θ_1 unaffected. We may note also that Jeffreys' independent prior $\pi^{JI}(\boldsymbol{\theta}) \propto \theta_3^{-1} \theta_4^{-1}$ is a first order probability matching prior, but is not a second order probability matching prior since it does not satisfy (2.9).

2.2.2 HPD Matching Prior

Writing $\mathbf{I}^{-1}(\boldsymbol{\theta}) = ((I^{ij}))$, due to orthogonality of θ_1 with $(\theta_2, \theta_3, \theta_4)$, from equation (33) of DiCiccio and Stern (1994) or equation (4.1) of Ghosh and Mukerjee (1995), the HPD equation in the present case reduces to

$$\frac{\partial^2}{\partial \theta_1^2} (I^{11} \pi) - \frac{\partial}{\partial \theta_1} \left((I^{11})^2 L_{111} \pi \right) - \sum_{\nu=2}^4 \frac{\partial}{\partial \theta_\nu} \left[\left(\sum_{s=2}^4 I^{\nu s} L_{11s} \right) I^{11} \pi \right] = 0 \quad , \quad (2.13)$$

where $L_{111} = E_{\boldsymbol{\theta}} \left(\frac{\partial^3 \log f}{\partial \theta_1^3} \right)$. In the present example, $L_{111} = 0$. Also, if π does not involve θ_1 , the first term in the left hand side of (2.13) is zero. With the other necessary substitutions, (2.13) simplifies in this case to

$$n_1^{-1} \frac{\partial}{\partial \theta_3} \left[\theta_3^3 \left(\theta_3^2/n_1 + \theta_4^2/n_2 \right)^{-1} \pi \right] + n_2^{-1} \frac{\partial}{\partial \theta_4} \left[\theta_4^3 \left(\theta_3^2/n_1 + \theta_4^2/n_2 \right)^{-1} \pi \right] = 0 \quad . \quad (2.14)$$

Clearly the prior $\pi^S(\boldsymbol{\theta})$ found in (2.11) satisfies (2.14), but $\pi^{JI}(\boldsymbol{\theta}) \propto (\theta_3\theta_4)^{-1}$ does not.

2.2.3 Matching via Inversion of Posterior Bartlett Corrected Likelihood Ratio Statistics

Writing $L_{1ij} = E \left[\frac{\partial^3 \log f}{\partial \theta_1 \partial \theta_i \partial \theta_j} \right]$, $2 \leq i, j \leq 4$, since in this example $L_{1ij} = 0$ for all $2 \leq i, j \leq 4$, the differential equation leading to the characterization of such priors is the same both for unconditional and conditional Bartlett adjusted LR statistic (*cf.* Yin and Ghosh, 1997, equations (2.3) and (2.4)). Further, since $L_{1,1,1} = 0$, it follows from Theorem 1 of Yin and Ghosh (1997) that the second order probability matching prior derived in (2.11) also has the property that the difference between Bayesian and frequentist Bartlett corrections for the conditional likelihood ratio statistic tends to zero as $N \rightarrow \infty$. Thus, the prior obtained (2.11) achieves matching in this way as well. In contrast, Jeffreys' independent prior $\pi^{JI}(\boldsymbol{\theta}) \propto \theta_3^{-1}\theta_4^{-1}$ does not satisfy either equation (2.3) or equation (2.4) of Yin and Ghosh (1997), and is not thus a matching prior under the criterion of this section.

2.2.4 Jeffreys' General Rule Prior

From (2.4), the determinant of $\mathbf{I}(\boldsymbol{\theta})$ is $|\mathbf{I}(\boldsymbol{\theta})| = 4n_1^2n_2^2\theta_3^{-4}\theta_4^{-4}$. Recall that Jeffreys' general rule prior π^{JG} is proportional to the square root of the determinant of the information matrix $\mathbf{I}(\boldsymbol{\theta})$. Hence

$$\pi^{JG}(\boldsymbol{\theta}) \propto \theta_3^{-2}\theta_4^{-2}$$

which is different from the Jeffreys' independent prior $\pi^{JI}(\boldsymbol{\theta})$ proposed by Jeffreys.

2.2.5 Reference Priors

In this section, we derive some reference priors for the Behrens-Fisher problem, keeping θ_1 fixed and making a one-to-one transformation of all the remaining parameters. We begin with the derivation of two group reference priors following the prescription of Berger and Bernardo (1989). Here θ_1 is the parameter of interest, and $(\theta_2, \theta_3, \theta_4)$ is the vector of nuisance parameters. The two-group reference prior is derived from the following four steps:

Step 1 $\pi(\theta_2, \theta_3, \theta_4 | \theta_1) = |\mathbf{I}_{22}(\boldsymbol{\theta})|^{1/2} = \theta_3^{-2} \theta_4^{-2} [4n_1^2 n_2^2 (\theta_3^2/n_1 + \theta_4^2/n_2)]^{1/2}$

Step 2 Consider the rectangular compacts $\Lambda_i = \{-i \leq \theta_1 \leq i, -i \leq \theta_2 \leq i, i^{-1} \leq \theta_3 \leq i, i^{-1} \leq \theta_4 \leq i\}$, $i = 1, 2, \dots$. Since $\int_{-i}^i \int_{i^{-1}}^i \int_{i^{-1}}^i \pi(\theta_2, \theta_3, \theta_4 | \theta_1) d\theta_2 d\theta_3 d\theta_4 = \text{constant}$, $\pi_i(\theta_2, \theta_3, \theta_4 | \theta_1) \propto \pi(\theta_2, \theta_3, \theta_4 | \theta_1) \propto \theta_3^{-2} \theta_4^{-2} (\theta_3^2/n_1 + \theta_4^2/n_2)^{1/2}$.

Step 3 $|\mathbf{I}(\boldsymbol{\theta})|/|\mathbf{I}_{22}(\boldsymbol{\theta})| = I_{11} = (\theta_3^2/n_1 + \theta_4^2/n_2)^{-1}$. Hence,

$$\pi_i(\theta_1) = \exp \left[\frac{1}{2} \int_{-i}^i \int_{i^{-1}}^i \int_{i^{-1}}^i \pi_i(\theta_2, \theta_3, \theta_4 | \theta_1) \log \frac{|\mathbf{I}(\boldsymbol{\theta})|}{|\mathbf{I}_{22}(\boldsymbol{\theta})|} d\theta_2 d\theta_3 d\theta_4 \right] = \text{constant}$$

Step 4 For any fixed point θ_{10} of θ_1

$$\lim_{i \rightarrow \infty} \left[\frac{K_i(\theta_1) \pi_i(\theta_1)}{K_i(\theta_{10}) \pi_i(\theta_{10})} \right] = 1, \quad \text{where } K_i(\theta_1) = 1 / \left[\int_{-i}^i \int_{i^{-1}}^i \int_{i^{-1}}^i \pi(\theta_2, \theta_3, \theta_4 | \theta_1) d\theta_2 d\theta_3 d\theta_4 \right].$$

Hence, from steps 1-4, the two-group reference prior is $\pi^{2G}(\boldsymbol{\theta}) \propto \theta_3^{-2} \theta_4^{-2} (\theta_3^2/n_1 + \theta_4^2/n_2)^{1/2}$ which is a first order probability matching prior, but is not either a second order probability matching prior, or a HPD matching prior or frequentist and Bayesian Bartlett adjusted matching prior. Similarly, one can find three group and one-at-a-time reference priors following the algorithm of Berger and Bernardo (1992b). Let $\boldsymbol{\theta}_{(1)} = \theta_1$, $\boldsymbol{\theta}_{(2)} = (\theta_3, \theta_4)$ and $\boldsymbol{\theta}_{(3)} = \theta_2$. Then, the Fisher information matrix for this grouping is given by

$$\mathbf{I}^*(\boldsymbol{\theta}) = \begin{bmatrix} (\theta_3^2/n_1 + \theta_4^2/n_2)^{-1} & \mathbf{0}^T \\ \mathbf{0} & \mathbf{I}_{22}^* \end{bmatrix}$$

where

$$\mathbf{I}_{22}^* = \begin{bmatrix} \frac{2n_1}{\theta_3^2} + \frac{4\theta_1^2\theta_4^2}{n_1n_2(\theta_3^2/n_1 + \theta_4^2/n_2)^3} & -\frac{4\theta_1^2\theta_3\theta_4}{n_1n_2(\theta_3^2/n_1 + \theta_4^2/n_2)^3} & \frac{2\theta_1}{\theta_3(\theta_3^2/n_1 + \theta_4^2/n_2)} \\ -\frac{4\theta_1^2\theta_3\theta_4}{n_1n_2(\theta_3^2/n_1 + \theta_4^2/n_2)^3} & \frac{2n_2}{\theta_4^2} + \frac{4\theta_1^2\theta_3^2}{n_1n_2(\theta_3^2/n_1 + \theta_4^2/n_2)^3} & -\frac{2\theta_1}{\theta_4(\theta_3^2/n_1 + \theta_4^2/n_2)} \\ \frac{2\theta_1}{\theta_3(\theta_3^2/n_1 + \theta_4^2/n_2)} & -\frac{2\theta_1}{\theta_4(\theta_3^2/n_1 + \theta_4^2/n_2)} & \frac{n_1}{\theta_3^2} + \frac{n_2}{\theta_4^2} \end{bmatrix} \quad (2.15)$$

and by Datta and Ghosh (1996)

$$|\mathbf{h}_1(\boldsymbol{\theta})| = \frac{|\mathbf{I}_{\sim 00}^*|}{|\mathbf{I}_{\sim 11}^*|} = \frac{n_1n_2}{n_2\theta_3^2 + n_1\theta_4^2}$$

$$|\mathbf{h}_2(\boldsymbol{\theta})| = \frac{|\mathbf{I}_{\sim 11}^*|}{|\mathbf{I}_{\sim 22}^*|} = \frac{4n_1n_2}{\theta_3^2\theta_4^2}$$

$$|\mathbf{h}_3(\boldsymbol{\theta})| = \frac{|\mathbf{I}_{\sim 22}^*|}{|\mathbf{I}_{\sim 33}^*|} = \frac{n_1}{\theta_3^2} + \frac{n_2}{\theta_4^2} \quad .$$

From the algorithm of Berger and Bernardo (1992b),

$$\begin{aligned} \pi_3^t(\boldsymbol{\theta}_{[\sim 2]}|\boldsymbol{\theta}_{[2]}) &= \pi_3^t(\boldsymbol{\theta}_{(3)}|\boldsymbol{\theta}_{[2]}) \\ &= \frac{|\mathbf{h}_3(\boldsymbol{\theta})|^{1/2}}{\int |\mathbf{h}_3(\boldsymbol{\theta})|^{1/2} d\boldsymbol{\theta}_{(3)}} \\ &\propto \text{constant} \\ \pi_2^t(\boldsymbol{\theta}_{[\sim 1]}|\boldsymbol{\theta}_{[1]}) &= \frac{\pi_3^t(\boldsymbol{\theta}_{[\sim 2]}|\boldsymbol{\theta}_{[2]}) \exp\left\{\frac{1}{2}E_2^t\left[(\log |\mathbf{h}_2(\boldsymbol{\theta})|)|\boldsymbol{\theta}_{[2]}\right]\right\}}{\int \int \exp\left\{\frac{1}{2}E_2^t\left[(\log |\mathbf{h}_2(\boldsymbol{\theta})|)|\boldsymbol{\theta}_{[2]}\right]\right\} d\boldsymbol{\theta}_{(2)}} \\ &\propto \theta_3^{-1}\theta_4^{-1} \\ \pi_1^t(\boldsymbol{\theta}_{[\sim 0]}|\boldsymbol{\theta}_{[0]}) &= \frac{\pi_2^t(\boldsymbol{\theta}_{[\sim 1]}|\boldsymbol{\theta}_{[1]}) \exp\left\{\frac{1}{2}E_1^t\left[(\log |\mathbf{h}_1(\boldsymbol{\theta})|)|\boldsymbol{\theta}_{[1]}\right]\right\}}{\int \int \exp\left\{\frac{1}{2}E_1^t\left[(\log |\mathbf{h}_1(\boldsymbol{\theta})|)|\boldsymbol{\theta}_{[1]}\right]\right\} d\boldsymbol{\theta}_{(1)}} \end{aligned}$$

$$\propto \pi_2^I(\boldsymbol{\theta}_{[\sim 1]}|\boldsymbol{\theta}_{[1]})$$

$$\propto \theta_3^{-1}\theta_4^{-1} \quad .$$

Hence, the three group reference prior for $\{\theta_1\}, \{\theta_3, \theta_4\}, \{\theta_2\}$ where the parameters are grouped according to their order of importance is $\pi^{3G}(\boldsymbol{\theta}) \propto \theta_3^{-1}\theta_4^{-1}$ which is Jeffreys' independent prior, but is different from the Jeffreys' general rule prior. Next, consider the one-at-a-time reference prior. Let $\boldsymbol{\theta}_{(1)} = \theta_1$, $\boldsymbol{\theta}_{(2)} = \theta_3$, $\boldsymbol{\theta}_{(3)} = \theta_4$ and $\boldsymbol{\theta}_{(3)} = \theta_2$. Then, the Fisher information matrix for this grouping is same as (2.15), and following Datta and Ghosh (1996), we get

$$|\mathbf{h}_1(\boldsymbol{\theta})| = \frac{|I_{[\sim 00]}^*|}{|I_{[\sim 11]}^*|} = \frac{n_1 n_2}{n_2 \theta_3^2 + n_1 \theta_4^2}$$

$$|\mathbf{h}_2(\boldsymbol{\theta})| = \frac{|I_{[\sim 11]}^*|}{|I_{[\sim 22]}^*|} = \frac{2n_1}{\theta_3^2}$$

$$|\mathbf{h}_3(\boldsymbol{\theta})| = \frac{|I_{[\sim 22]}^*|}{|I_{[\sim 33]}^*|} = \frac{2n_2}{\theta_4^2}$$

$$|\mathbf{h}_4(\boldsymbol{\theta})| = \frac{|I_{[\sim 33]}^*|}{|I_{[\sim 44]}^*|} = \frac{n_1}{\theta_3^2} + \frac{n_2}{\theta_4^2} \quad .$$

Also, from the algorithm of Berger and Bernardo (1992b),

$$\begin{aligned} \pi_4^I(\boldsymbol{\theta}_{[\sim 3]}|\boldsymbol{\theta}_{[3]}) &= \pi_4^I(\boldsymbol{\theta}_{(4)}|\boldsymbol{\theta}_{[3]}) \\ &= \frac{|\mathbf{h}_4(\boldsymbol{\theta})|^{1/2}}{\int |\mathbf{h}_4(\boldsymbol{\theta})|^{1/2} d\boldsymbol{\theta}_{(4)}} \end{aligned}$$

$$\propto \text{constant}$$

$$\pi_3^I(\boldsymbol{\theta}_{[\sim 2]}|\boldsymbol{\theta}_{[2]}) = \frac{\pi_4^I(\boldsymbol{\theta}_{[\sim 3]}|\boldsymbol{\theta}_{[3]}) \exp \left\{ \frac{1}{2} E_3^I \left[(\log |\mathbf{h}_3(\boldsymbol{\theta})|) |\boldsymbol{\theta}_{[3]}] \right] \right\}}{\int \exp \left\{ \frac{1}{2} E_3^I \left[(\log |\mathbf{h}_3(\boldsymbol{\theta})|) |\boldsymbol{\theta}_{[3]}] \right] \right\} d\boldsymbol{\theta}_{(2)}}$$

$$\begin{aligned}
& \propto \theta_4^{-1} \\
\pi_2^l(\theta_{[\sim 1]}|\theta_{[1]}) &= \frac{\pi_3^l(\theta_{[\sim 2]}|\theta_{[2]}) \exp\left\{\frac{1}{2}E_2^l\left[(\log|\mathbf{h}_2(\boldsymbol{\theta})|)|\theta_{[2]}\right]\right\}}{\int \exp\left\{\frac{1}{2}E_2^l\left[(\log|\mathbf{h}_3(\boldsymbol{\theta})|)|\theta_{[2]}\right]\right\} d\boldsymbol{\theta}_{(1)}} \\
& \propto \theta_3^{-1}\theta_4^{-1} \\
\pi_1^l(\theta_{[\sim 0]}|\theta_{[0]}) &= \frac{\pi_2^l(\theta_{[\sim 1]}|\theta_{[1]}) \exp\left\{\frac{1}{2}E_1^l\left[(\log|\mathbf{h}_1(\boldsymbol{\theta})|)|\theta_{[1]}\right]\right\}}{\int \int \exp\left\{\frac{1}{2}E_1^l\left[(\log|\mathbf{h}_1(\boldsymbol{\theta})|)|\theta_{[1]}\right]\right\} d\boldsymbol{\theta}_{(1)}} \\
& \propto \pi_2^l(\theta_{[\sim 1]}|\theta_{[1]}) \\
& \propto \theta_3^{-1}\theta_4^{-1} \quad .
\end{aligned}$$

Under the groupings $\{\theta_1\}, \{\theta_4\}, \{\theta_3\}, \{\theta_2\}$, the same prior is obtained as one-at-a-time reference priors. Other groupings are possible, but these do not yield analytically tractable reference priors. The catalog of reference priors is given in Table 2.1.

Table 2.1: Catalog of reference priors for Behrens-Fisher Problem

Grouped parameters in their order of importance	Prior distribution
$\{\theta_1\}, \{\theta_2, \theta_3, \theta_4\}$	$\pi^{2G}(\boldsymbol{\theta}) \propto \theta_3^{-2}\theta_4^{-2}(\theta_3^2/\rho_1 + \theta_4^2/\rho_2)^{1/2}$
$\{\theta_1\}, \{\theta_3, \theta_4\}, \{\theta_2\}$	$\pi^{3G}(\boldsymbol{\theta}) \propto \theta_3^{-1}\theta_4^{-1}$
$\{\theta_1\}, \{\theta_3\}, \{\theta_4\}, \{\theta_2\}$	$\pi^{1G}(\boldsymbol{\theta}) \propto \theta_3^{-1}\theta_4^{-1}$
$\{\theta_1\}, \{\theta_4\}, \{\theta_3\}, \{\theta_2\}$	$\pi^{1G}(\boldsymbol{\theta}) \propto \theta_3^{-1}\theta_4^{-1}$

2.3 Propriety of Posteriors and Simulation Study

As we mentioned in previous section, Jeffreys' independent prior π^{JI} is the same as some reference priors such as π^{3G} and π^{1G} . We denote this prior as π^R . The first task is to verify the propriety of posteriors under the noninformative priors π^R , π^{JG} and π^S . To this end, we find it convenient to go back to the original parameterization $(\mu_1, \mu_2, \sigma_1, \sigma_2)$. The joint posterior of μ_1, μ_2, σ_1 and σ_2 given $\mathbf{x}_1, \mathbf{x}_2$ under the prior π^R , π^{JG} and π^S are given respectively

$$\begin{aligned} \pi^R(\mu_1, \mu_2, \sigma_1, \sigma_2 | \mathbf{x}_1, \mathbf{x}_2) &\propto \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{j=1}^{n_1} (x_{1j} - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{j=1}^{n_2} (x_{2j} - \mu_2)^2 \right\} \\ &\times \sigma_1^{-(n_1+1)} \sigma_2^{-(n_2+1)} \quad , \end{aligned} \quad (2.16)$$

$$\begin{aligned} \pi^{JG}(\mu_1, \mu_2, \sigma_1, \sigma_2 | \mathbf{x}_1, \mathbf{x}_2) &\propto \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{j=1}^{n_1} (x_{1j} - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{j=1}^{n_2} (x_{2j} - \mu_2)^2 \right\} \\ &\times \sigma_1^{-(n_1+2)} \sigma_2^{-(n_2+2)} \quad , \end{aligned} \quad (2.17)$$

$$\begin{aligned} \pi^S(\mu_1, \mu_2, \sigma_1, \sigma_2 | \mathbf{x}_1, \mathbf{x}_2) &\propto \exp \left\{ -\frac{1}{2\sigma_1^2} \sum_{j=1}^{n_1} (x_{1j} - \mu_1)^2 - \frac{1}{2\sigma_2^2} \sum_{j=1}^{n_2} (x_{2j} - \mu_2)^2 \right\} \\ &\times \sigma_1^{-(n_1+3)} \sigma_2^{-(n_2+3)} (\sigma_1^2/n_1 + \sigma_2^2/n_2) \quad . \end{aligned} \quad (2.18)$$

From (2.16), integration with respect to σ_1 and σ_2 yields

$$\pi^R(\mu_1, \mu_2 | \mathbf{x}_1, \mathbf{x}_2) \propto \left[1 + \frac{n_1(\mu_1 - \bar{x}_1)^2}{(n_1 - 1)s_1^2} \right]^{-\frac{n_1}{2}} \left[1 + \frac{n_2(\mu_2 - \bar{x}_2)^2}{(n_2 - 1)s_2^2} \right]^{-\frac{n_2}{2}} . \quad (2.19)$$

Thus, the joint posterior of μ_1 and μ_2 based on the prior π^R is a product of two independent t 's with respective location parameters $\bar{x}_1, \bar{x}_2$, scale parameters $s_1/\sqrt{n_1}$ and $s_2/\sqrt{n_2}$ and degrees of freedom $n_1 - 1$ and $n_2 - 1$. This posterior is proper for $n_1 \geq 2$ and $n_2 \geq 2$.

Similarly, from (2.17), integration with respect to σ_1 and σ_2 yields

$$\pi^{JG}(\mu_1, \mu_2 | \mathbf{x}_1, \mathbf{x}_2) \propto \left[1 + \frac{n_1(\mu_1 - \bar{x}_1)^2}{(n_1 - 1)s_1^2} \right]^{-\frac{n_1+1}{2}} \left[1 + \frac{n_2(\mu_2 - \bar{x}_2)^2}{(n_2 - 1)s_2^2} \right]^{-\frac{n_2+1}{2}}. \quad (2.20)$$

Thus, the joint posterior of μ_1 and μ_2 based on the prior π^{JG} is a product of two independent t 's with respective location parameters $\bar{x}_1, \bar{x}_2$, scale parameters $s_1/\sqrt{n_1}$ and $s_2/\sqrt{n_2}$ and degrees of freedom n_1 and n_2 . This posterior is proper for $n_1 \geq 1$ and $n_2 \geq 1$

Next, from (2.18), integration with respect to σ_1 and σ_2 yields

$$\begin{aligned} \pi^S(\mu_1, \mu_2 | \mathbf{x}_1, \mathbf{x}_2) &\propto C n_1^{-1} \left[\sum_{j=1}^{n_1} (x_{1j} - \mu_1)^2 \right]^{-\frac{n_1}{2}} \left[\sum_{j=1}^{n_2} (x_{2j} - \mu_2)^2 \right]^{-\frac{n_2}{2}} \\ &\quad + C n_2^{-1} \left[\sum_{j=1}^{n_1} (x_{1j} - \mu_1)^2 \right]^{-\frac{n_1+2}{2}} \left[\sum_{j=1}^{n_2} (x_{2j} - \mu_2)^2 \right]^{-\frac{n_2}{2}} \end{aligned} \quad (2.21)$$

where the constant C depends on $\mathbf{x}_1, \mathbf{x}_2, n_1$ and n_2 , but not on the parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2)$. We rewrite the right hand side of (2.21) as

$$\begin{aligned} &C \left[n_1^{-1} \sqrt{\frac{u_1}{n_1}} B\left(\frac{1}{2}, \frac{n_1-1}{2}\right) u_1^{-\frac{n_1}{2}} \left[1 + \frac{n_1(\mu_1 - \bar{x}_1)^2}{(n_1-1)(u_1/(n_1-1))} \right]^{-\frac{n_1}{2}} \frac{1}{\sqrt{n_1-1} B(\frac{1}{2}, \frac{n_1-1}{2})} \sqrt{\frac{n_1(n_1-1)}{u_1}} \right. \\ &\quad \times \sqrt{\frac{u_2}{n_2}} B\left(\frac{1}{2}, \frac{n_2+1}{2}\right) u_2^{-\frac{n_2+2}{2}} \left[1 + \frac{n_2(\mu_2 - \bar{x}_2)^2}{(n_2+1)(u_2/(n_2+1))} \right]^{-\frac{n_2+2}{2}} \frac{1}{\sqrt{n_2+1} B(\frac{1}{2}, \frac{n_2+1}{2})} \sqrt{\frac{n_2(n_2+1)}{u_2}} \\ &\quad + n_2^{-1} \sqrt{\frac{u_1}{n_1}} B\left(\frac{1}{2}, \frac{n_1+1}{2}\right) u_1^{-\frac{n_1+2}{2}} \left[1 + \frac{n_1(\mu_1 - \bar{x}_1)^2}{(n_1+1)(u_1/(n_1+1))} \right]^{-\frac{n_1+2}{2}} \frac{1}{\sqrt{n_1+1} B(\frac{1}{2}, \frac{n_1+1}{2})} \sqrt{\frac{n_1(n_1+1)}{u_1}} \\ &\quad \left. \times \sqrt{\frac{u_2}{n_2}} B\left(\frac{1}{2}, \frac{n_2-1}{2}\right) u_2^{-\frac{n_2}{2}} \left[1 + \frac{n_2(\mu_2 - \bar{x}_2)^2}{(n_2-1)(u_2/(n_2-1))} \right]^{-\frac{n_2}{2}} \frac{1}{\sqrt{n_2-1} B(\frac{1}{2}, \frac{n_2-1}{2})} \sqrt{\frac{n_2(n_2-1)}{u_2}} \right] \end{aligned} \quad (2.22)$$

where $u_i = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$, ($i = 1, 2$) and $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$. Now, after some simplifications, one can write

$$\pi^S(\mu_1, \mu_2 | \mathbf{x}_1, \mathbf{x}_2) = \omega f_{\bar{x}_1, u_1^{1/2}(n_1(n_1-1))^{-\frac{1}{2}}, n_1-1}(\mu_1) f_{\bar{x}_2, u_2^{1/2}(n_2(n_2+1))^{-\frac{1}{2}}, n_2+1}(\mu_2)$$

$$+ (1 - \omega) f_{\bar{x}_1, u_1^{1/2}(n_1(n_1+1))^{-\frac{1}{2}}, n_1+1}(\mu_1) f_{\bar{x}_2, u_2^{1/2}(n_2(n_2-1))^{-\frac{1}{2}}, n_2-1}(\mu_2) \quad (2.23)$$

where $f_{\phi_1, \phi_2, \nu}(\cdot)$ denotes the pdf of Student's t with location parameter ϕ_1 , scale parameter ϕ_2 and degrees of freedom ν , and

$$\omega = \frac{u_1(n_1 - 1)^{-1}}{u_1(n_1 - 1)^{-1} + u_2(n_2 - 1)^{-1}} = \frac{s_1^2}{s_1^2 + s_2^2}.$$

The propriety of the posterior $\pi^S(\mu_1, \mu_2, \sigma_1, \sigma_2 | \mathbf{x}_1, \mathbf{x}_2)$ follows immediately from (2.19) and (2.23) for $n_1 \geq 2$ and $n_2 \geq 2$.

Next we prove a theorem which establishes the symmetry and unimodality of the posterior of $\mu_1 - \mu_2$ around $\bar{x}_1 - \bar{x}_2$ under the priors π^R , π^{JG} and π^S .

Theorem 2.1 The posterior of $\mu_1 - \mu_2$ is symmetric and unimodal about $\bar{x}_1 - \bar{x}_2$ under the priors π^R , π^{JG} and π^S .

Proof of Theorem 2.1 From (2.19), (2.20) and (2.23), it is easy to check that (μ_1, μ_2) is symmetric about $(\bar{x}_1, \bar{x}_2)$ under the priors π^R , π^{JG} and π^S . This implies that $\mu_1 - \mu_2$ is symmetric about $\bar{x}_1 - \bar{x}_2$ under these priors. In order to prove unimodality, first under the priors π^R and π^{JG} , we observe that the posterior pdf's $\mu_1 - \bar{x}_1$ and $\mu_2 - \bar{x}_2$ are both symmetric and unimodal about 0 on the real line. Since the concepts of symmetric unimodality and central convex unimodality coincide on the real line, both these posteriors are central convex unimodal about 0. Due to the independence of these two posteriors, from Theorem 2.4, pp 45-46 of Dharmadhikari and Joag-dev (1988), the joint posterior of $(\mu_1 - \bar{x}_1, \mu_2 - \bar{x}_2)$ is unimodal about $(0, 0)$. Now, from Theorem 2.15, p 59 of Dharmadhikari and Joagdev (1988), central convex unimodality implies linear unimodality, and hence $(\mu_1 - \mu_2) - (\bar{x}_1 - \bar{x}_2)$ is unimodal about 0, or equivalently $\mu_1 - \mu_2$ is unimodal about $\bar{x}_1 - \bar{x}_2$.

In order to prove the unimodality of $\mu_1 - \mu_2$ about $\bar{x}_1 - \bar{x}_2$ under the prior π^S , first, as before, the joint posteriors $f_{\bar{x}_1, u_1^{1/2}(n_1(n_1-1))^{-\frac{1}{2}}, n_1-1}(\mu_1) f_{\bar{x}_2, u_2^{1/2}(n_2(n_2+1))^{-\frac{1}{2}}, n_2+1}(\mu_2)$

and $f_{\bar{x}_1, u_1^{1/2}(n_1(n_1+1))^{-\frac{1}{2}}, n_1+1}(\mu_1) f_{\bar{x}_2, u_2^{1/2}(n_2(n_2-1))^{-\frac{1}{2}}, n_2-1}(\mu_2)$ are both central convex unimodal about $(\bar{x}_1, \bar{x}_2)$. Hence, the convex combination (2.23) is also central convex unimodal about $(\bar{x}_1, \bar{x}_2)$. The same Theorem 2.15 of Dharmadhikari and Joag-dev now establishes unimodality of $\mu_1 - \mu_2$ about $\bar{x}_1 - \bar{x}_2$.

Next we undertake some simulation study to find the frequentist coverage probabilities of the Bayesian credible intervals for $\mu_1 - \mu_2$ based on the priors π^R , π^{JG} and π^S . We first generate samples $\mathbf{x}_1, \mathbf{x}_2$ of sizes n_1 and n_2 for a given choice of $(\mu_1, \mu_2, \sigma_1, \sigma_2)$. A variety of (n_1, n_2) combinations are considered, in some instances, n_1 and n_2 are close, while on other occasions, they are quite dispersed. Then, from the joint posterior derived in (2.19), (2.20) and (2.23), we generate samples 1,000 times, compute $\mu_1 - \mu_2$ each time, and find numerically the 5% and 95% posterior quantiles of $\mu_1 - \mu_2$. Due to the symmetry and unimodality of the posterior pdf's of $\mu_1 - \mu_2$ about $\bar{x}_1 - \bar{x}_2$ under the priors π^R , π^{JG} and π^S , we consider equal tailed credible intervals and check whether or not the actual difference $\mu_1 - \mu_2$ belongs to this interval. The whole process is repeated 5,000 times, and we find the proportion of times the true difference $\mu_1 - \mu_2$ belongs to this interval. This is the estimated frequentist coverage probability of the Bayesian credible interval. The results are given in Tables 2.2 and 2.3, respectively. The computing was done on Sun Sparc10 workstation using interface FORTRAN77. All random numbers are generated by FORTRAN77 subroutine based on *Uniform*(0, 1) random variable.

The findings of Tables 2.2 and 2.3 are quite impressive for the prior π^S when compared to the priors π^R and π^{JG} especially for small and moderate values of n_1 and n_2 . For the cases $n_1 = 2$ and $n_2 = 20$ or $n_1 = 20$ and $n_2 = 2$, the coverage probabilities of the Bayesian credible intervals based on the second order probability matching prior meet the frequentist target (here 90%) better than Jeffreys' independent (or the one-at-a-time reference) prior. Also, for all the cases considered here,

Table 2.2: Frequentist probabilities of 0.05 (0.95) posterior quantiles of θ_1 when $\mu_1 = 2.0, \mu_2 = 0.0$ and $\sigma_2^2 = 1.0$ ($n_1 > n_2$)

n_1	n_2	π	$\sigma_1^2 = 2.0$		$\sigma_1^2 = 4.0$	
			0.05	0.95	0.05	0.95
2	2	π^R	0.007	0.991	0.010	0.989
		π^{JG}	0.085	0.915	0.086	0.913
		π^S	0.017	0.979	0.024	0.979
20	2	π^R	0.036	0.803	0.029	0.822
		π^{JG}	0.100	0.902	0.077	0.922
		π^S	0.110	0.856	0.086	0.878
3	2	π^R	0.013	0.988	0.011	0.987
		π^{JG}	0.059	0.928	0.064	0.928
		π^S	0.028	0.970	0.030	0.970
5	3	π^R	0.021	0.976	0.025	0.978
		π^{JG}	0.054	0.945	0.054	0.944
		π^S	0.038	0.958	0.042	0.960
7	5	π^R	0.031	0.969	0.034	0.959
		π^{JG}	0.047	0.949	0.051	0.943
		π^S	0.041	0.955	0.045	0.948
12	8	π^R	0.041	0.962	0.043	0.956
		π^{JG}	0.051	0.952	0.053	0.948
		π^S	0.051	0.952	0.050	0.950
15	10	π^R	0.043	0.960	0.048	0.956
		π^{JG}	0.051	0.950	0.056	0.948
		π^S	0.051	0.954	0.053	0.952
20	15	π^R	0.046	0.958	0.044	0.955
		π^{JG}	0.051	0.953	0.048	0.951
		π^S	0.052	0.953	0.049	0.951
30	20	π^R	0.049	0.951	0.047	0.953
		π^{JG}	0.051	0.945	0.051	0.950
		π^S	0.051	0.948	0.051	0.950

Table 2.3: Frequentist coverage probabilities of 0.05 (0.95) posterior quantiles of θ_1 when $\mu_1 = 2.0, \mu_2 = 0.0$ and $\sigma_2^2 = 1.0$ ($n_1 < n_2$)

n_1	n_2	π	$\sigma_1^2 = 2.0$		$\sigma_1^2 = 4.0$	
			0.05	0.95	0.05	0.95
2	3	π^R	0.015	0.983	0.023	0.977
		π^{JG}	0.085	0.919	0.092	0.901
		π^S	0.036	0.965	0.042	0.952
2	20	π^R	0.036	0.788	0.043	0.788
		π^{JG}	0.120	0.880	0.123	0.861
		π^S	0.110	0.819	0.104	0.819
3	5	π^R	0.033	0.968	0.033	0.964
		π^{JG}	0.068	0.933	0.074	0.924
		π^S	0.050	0.950	0.050	0.945
5	7	π^R	0.039	0.963	0.045	0.962
		π^{JG}	0.056	0.943	0.064	0.940
		π^S	0.049	0.952	0.052	0.952
8	12	π^R	0.046	0.953	0.044	0.956
		π^{JG}	0.058	0.941	0.057	0.944
		π^S	0.055	0.945	0.051	0.949
10	15	π^R	0.044	0.952	0.044	0.949
		π^{JG}	0.052	0.944	0.053	0.943
		π^S	0.051	0.948	0.048	0.948
15	20	π^R	0.046	0.955	0.047	0.953
		π^{JG}	0.054	0.949	0.052	0.947
		π^S	0.053	0.951	0.049	0.952
20	30	π^R	0.048	0.953	0.048	0.952
		π^{JG}	0.052	0.950	0.051	0.946
		π^S	0.051	0.950	0.049	0.951

the coverage probability of the one-sided credible interval based on the second order probability matching prior is closer to the frequentist target at the upper 5% level. The same is true for the lower 5% level as well except for the cases $n_1 = 2$ and $n_2 = 20$ or $n_1 = 20$ and $n_2 = 2$. More important, there is noticeable differences between the two approaches for small and moderate values of n_1 and n_2 . Finally, one does not require very large n_1 and n_2 for these priors to be useful.

2.4 Conditional Frequentist Properties

As mentioned in the introduction, the Welch-Aspin solution has been criticized by Robinson (1982) because this may lead to a negatively biased relevant subset in the sense of Buehler (1959). To be specific, if $\mathbf{I}_{WA}(\mathbf{X}_1, \mathbf{X}_2)$ denotes the Welch-Aspin intervals for $\mu_1 - \mu_2$, then there exists some subset $\mathcal{C}$ of the sample space such that

$$P[\mu_1 - \mu_2 \in \mathbf{I}_{WA}(\mathbf{X}_1, \mathbf{X}_2) | (\mathbf{X}_1, \mathbf{X}_2) \in \mathcal{C}] \leq 1 - \alpha - \epsilon, \quad (0 < \epsilon < 1 - \alpha) \quad (2.24)$$

for all μ_1, μ_2, σ_1 and σ_2 . On the other hand, Robinson (1976) showed that credible intervals based on Jeffreys' independent prior π^{JI} does not admit any negatively biased relevant set in the sense of (2.24). Since the credible interval based on Jeffreys' independent prior is algebraically equivalent to the original Behrens-Fisher solution, it follows that the latter is more satisfactory than the Welch-Aspin solution from a conditional frequentist perspective since the conditional coverage probabilities of the Behrens-Fisher intervals are never smaller than the nominal confidence level.

The alternate prior π^S also does not admit any negatively biased relevant set. This can be proved by mimicking the arguments of Robinson (1976) and by modifying appropriately the sequence of priors considered by him. The main point is that any Bayesian credible interval for $\mu_1 - \mu_2$ usually has this nice conditional property as long as it contains $\bar{x}_1 - \bar{x}_2$.

2.5 Concluding Remarks

In this chapter, we have derived a new prior different from Jeffreys' independent and general rule priors for the Behrens-Fisher problem. This prior possesses good frequentist properties in that the coverage probabilities of credible intervals for the difference of two normal means based on this prior match their frequentist counterpart very closely even for small and moderate sample sizes. The proposed prior is also very satisfactory from a conditional frequentist perspective.

Default priors are being routinely used by Bayesians (and often by non-Bayesians as well) to analyze real life data. We do not expect that there will be a single default prior which will stand out well above the others on every single occasion. Based on the four criteria as mentioned in the introduction, there could be different default priors each optimal according to one of these criteria. On occasions, the one-at-a-time reference prior emerges as the optimal one both according to criteria (i) and (ii). Such examples are found in Tibshirani (1989), Mukerjee and Dey (1993) and Datta and Ghosh, M. (1995). In these instances, clearly one should prescribe the one-at-a-time reference prior as the default prior.

The situation is not so clear-cut for the Behrens-Fisher problem. It is well-known in this case that the Welch-Aspin frequentist solution is not satisfactory even from a conditional frequentist perspective, while Behrens' solution, though algebraically equivalent to the solution provided by Jeffreys' independent prior, has not found appeal to frequentists. One major motivation behind the present work was to find (if possible) a prior which should be of appeal to frequentists, conditional frequentists as well as default Bayesians. It seems to us that the proposed second order matching prior meets this end quite well.

Clearly there are possible extensions of the proposed work. One possible extension is for the ANOVA model where homoscedasticity fails, and one is interested in finding a suitable confidence interval for a treatment contrast. A different but related problem

is to find a confidence interval for the common mean of several independent normal populations with possibly unequal variances. Development of default priors in these cases should be of appeal to both Bayesians and frequentists. We plan to address some of these problems in later studies.

CHAPTER 3

INTERVAL ESTIMATION OF THE COMMON MEAN OF SEVERAL NORMAL POPULATIONS

3.1 Introduction

Estimation of the common mean of several normal populations has received attention for several decades. The problem is quite natural in balanced incomplete block designs with uncorrelated random block effects and fixed treatment effects $\hat{\tau}$ and the interblock estimator $\tilde{\tau}$ of a treatment contrast are independent normal with a common mean τ , but their variances are unknown and unequal. A second example relates to meta analysis where for example, several clinics provide estimates of a common parameter of interest, and the problem is how to combine these estimates meaningfully into a single one.

Point estimation of the common mean has been addressed both from the classical and decision theoretic points of view. Among others, we may refer to Graybill and Deal (1959), Zacks (1966,1970), Khatri and Shah (1974), Shinozaki (1978), Cohen and Sackrowitz (1974), Brown and Cohen (1974), Bhattacharya (1980), Sinha and Mouqadem (1982), Sinha (1985) and Kubokawa (1990).

In contrast, relatively less attention has been paid to interval estimation of the common mean. Approximate confidence intervals are found in Meier (1953), Brown and Cohen (1974), Sinha (1985) and Eberhardt, Reeve and Spiegelman (1989). Exact intervals are proposed in Fairweather (1972), Cohen and Sackrowitz (1974) and Jordan and Krishnamoorthy (1996). More recently, Sinha (1998) and Yu, Sun and Sinha (1999) have these intervals and others exact intervals compared by their lengths when the confidence coefficients remain the same.

From the examples considered in Yu, Sun and Sinha (1999), it appears that none of the frequentist procedures emerges as a clear winner. Indeed, in one of the two examples considered in their paper, the Fairweather interval is the shortest, while in the second example the inverse normal interval is the shortest. Thus, a unified frequentist method for finding an exact confidence interval of the common mean seems to be lacking.

In this chapter, we offer an alternative approach by introducing Bayesian priors such that the resulting credible intervals for the common mean have coverage probabilities asymptotically equivalent to their frequentist counterparts. Though asymptotic, our simulation results indicate that this matching holds even for small and moderate sample sizes. Also, for both the examples considered by Yu, Sun and Sinha (1999), the lengths of the Bayesian intervals are nearly equivalent to the shortest frequentist intervals among those considered in Yu, Sun and Sinha (1999). While the appeal of a Bayesian procedure is not just by judging its frequentist property, in the absence of prior elicitation, one is somewhat compelled to use default priors, and one criterion for choosing such priors is by judging their frequentist properties.

The outline of the remaining sections is as follows. In Section 2, we introduce several criteria of matching priors, and obtain one which stands out particularly well under any of these criteria. Propriety of the posterior under this prior is proved in Section 3. A small simulation study is undertaken in Section 4 which shows that the second order probability matching prior developed in this chapter matches the frequentist target coverage probabilities better than the reference prior especially for small or moderate samples. Section 5 contains several confidence intervals that are proposed from a frequentist point of view. Also, in this section, the length of the Bayesian credible interval is compared against all its frequentist competitors for two real-life examples considered in Yu, Sun and Sinha (1999).

3.2 Noninformative Priors

Suppose there are $k(\geq 2)$ independent normal populations with unknown common mean μ and with unknown and possibly unequal variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$. We consider samples of sizes $n_i = N\rho_i$ ($i = 1, \dots, k$), where each $n_i \geq 2$ and $\sum_{i=1}^k \rho_i = 1$. The sample observations are denoted by X_{ij} ($j = 1, \dots, n_i; i = 1, \dots, k$). The minimal sufficient statistic is $(\bar{X}_1, \dots, \bar{X}_k, S_1^2, \dots, S_k^2)$ where $\bar{X}_i = n_i^{-1} \sum_{j=1}^{n_i} X_{ij}$ and $S_i^2 = (n_i - 1)^{-1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$. The likelihood function is given by

$$L(\mu, \sigma_1, \dots, \sigma_k) \propto \left(\prod_{i=1}^k \sigma_i^{-n_i} \right) \exp \left[-\frac{1}{2} \sum_{i=1}^k \sigma_i^{-2} \left\{ n_i (\mu - \bar{X}_i)^2 + (n_i - 1) S_i^2 \right\} \right]. \quad (3.1)$$

Simple calculations yield the Fisher information matrix as

$$I(\mu, \sigma_1, \dots, \sigma_k) = \text{Diagonal} \left(\sum_{i=1}^k n_i \sigma_i^{-2}, 2n_1 \sigma_1^{-2}, \dots, 2n_k \sigma_k^{-2} \right). \quad (3.2)$$

We now consider several noninformative priors all having their roots in the Fisher information matrix given in (3.2) in some form or the other.

3.2.1 Jeffreys' Prior

Jeffreys' prior π^J is proportional to the positive square root of the determinant of the $I(\mu, \sigma_1, \dots, \sigma_k)$. Thus,

$$\pi^J(\mu, \sigma_1, \dots, \sigma_k) \propto \left(\sum_{i=1}^k n_i \sigma_i^{-2} \right)^{1/2} \left(\prod_{i=1}^k \sigma_i^{-1} \right). \quad (3.3)$$

3.2.2 Reference Priors

First we consider the two-group reference prior. The parameter space is partitioned into $\{\mu\}, \{\sigma_1, \dots, \sigma_k\}$, μ being the parameter of interest. Let $\theta_{(1)} = \mu$ and $\theta_{(2)} = (\sigma_1, \dots, \sigma_k)$. Then, the Fisher information matrix is written by

$$\mathbf{I}(\theta) = \text{Block Diagonal} \left(h_1(\theta), h_2(\theta) \right) ,$$

where

$$\begin{aligned} h_1(\theta) &= \sum_{i=1}^k n_i \sigma_i^{-2} \\ h_2(\theta) &= \text{Diagonal} \left(2n_1 \sigma_1^{-2}, \dots, 2n_k \sigma_k^{-2} \right) . \end{aligned}$$

Hence, from Theorem 1 of Datta and M.Ghosh (1995), the two-group reference prior is given by

$$\pi^{2R}(\mu, \sigma_1, \dots, \sigma_k) \propto \prod_{i=1}^k \sigma_i^{-1} . \quad (3.4)$$

Next consider the one-at-a-time reference prior with $\theta = (\theta_1, \dots, \theta_p) = (\mu, \sigma_1, \dots, \sigma_k)$. Then, from the Lemma of Datta and Ghosh (1996),

$$\begin{aligned} |h_1(\theta)| &= \frac{|\mathbf{I}_{[\sim 00]}|}{|\mathbf{I}_{[\sim 11]}|} = \sum_{i=1}^k n_i \sigma_i^{-2} \\ |h_2(\theta)| &= \frac{|\mathbf{I}_{[\sim 11]}|}{|\mathbf{I}_{[\sim 22]}|} = \frac{2n_1}{\sigma_1^2} \\ &\vdots \\ |h_k(\theta)| &= \frac{|\mathbf{I}_{[\sim k-1, k-1]}|}{|\mathbf{I}_{[\sim k k]}|} = \frac{2n_{k-1}}{\sigma_{k-1}^2} \\ |h_{k+1}(\theta)| &= \frac{|\mathbf{I}_{[\sim k k]}|}{|\mathbf{I}_{[\sim k+1, k+1]}|} = \frac{2n_k}{\sigma_k^2} . \end{aligned}$$

Also, following the algorithm of Berger and Bernardo (1992b) and Datta and Ghosh (1996) as the previous chapter, the one-at-a-time reference prior for $\{\mu\}, \{\sigma_1\}, \dots, \{\sigma_k\}$ is

$$\pi^R(\mu, \sigma_1, \dots, \sigma_k) \propto \prod_{i=1}^k \sigma_i^{-1} \quad (3.5)$$

3.2.3 Quantile Matching Priors

Due to the orthogonality of μ with $(\sigma_1, \dots, \sigma_k)$ (*cf.* Cox and Reid, 1987), from Tibshirani (1989), the class of first order probability matching priors is characterized by

$$\pi^F(\mu, \sigma_1, \dots, \sigma_k) \propto \left(\sum_{i=1}^k n_i \sigma_i^{-2} \right)^{1/2} g(\sigma_1, \dots, \sigma_k), \quad (3.6)$$

where g is any arbitrary positive-valued function differentiable in its arguments. Thus both Jeffreys' prior and the one-at-a-time (or two-group) reference priors are first order probability matching priors. As one may note, from (3.6), any smooth function of $\sigma_1, \dots, \sigma_k$ differentiable in its arguments is a first order probability matching prior. This class of priors, however, is too large to be useful in practice. Hence, there is a need to narrow down the selection of priors within this class, and in particular, develop second order probability matching priors.

From Mukerjee and Ghosh (1997), due to the orthogonality of μ with $(\sigma_1, \dots, \sigma_k)$, this class of priors is characterized as solutions of the second order partial differential equation

$$\frac{1}{6} g(\sigma_1, \dots, \sigma_k) \frac{\partial}{\partial \mu} (I_{11}^{-3/2} L_{1,1,1}) + \sum_{i=1}^k \frac{\partial}{\partial \sigma_i} \{ I_{11}^{-1/2} L_{11i} (\sigma_i^2 / 2n_i) g(\sigma_1, \dots, \sigma_k) \} = 0, \quad (3.7)$$

where $L_{1,1,1} = E \left[\left(\frac{\partial \log L}{\partial \mu} \right)^3 \middle| \mu, \sigma_1, \dots, \sigma_k \right]$ and $L_{11i} = E \left[\frac{\partial^3 \log L}{\partial \mu^2 \partial \sigma_i} \middle| \mu, \sigma_1, \dots, \sigma_k \right]$, $i = 1, \dots, k$.

From (3.1), after some algebra, we have

$$L_{1,1,1} = 0$$

$$L_{11i} = \frac{2n_i}{\sigma_i^3}, \quad i = 1, \dots, k \quad .$$

Hence, (3.7) simplifies to

$$\sum_{i=1}^k \frac{\partial}{\partial \sigma_i} \left\{ I_{11}^{-1/2} \sigma_i^{-1} g(\sigma_1, \dots, \sigma_k) \right\} = 0 \quad . \quad (3.8)$$

A general solution to (3.8) is difficult to find. However, one particular solution is given by $g(\sigma_1, \dots, \sigma_k) \propto I_{11}^{1/2} \left(\prod_{i=1}^k \sigma_i \right)$. Accordingly, one particular second order probability matching prior is given by

$$\begin{aligned} \pi^S(\mu, \sigma_1, \dots, \sigma_k) &\propto \left(\prod_{i=1}^k \sigma_i \right) I_{11} \\ &= \left(\prod_{i=1}^k \sigma_i \right) \left(\sum_{i=1}^k n_i \sigma_i^{-2} \right) \quad . \end{aligned} \quad (3.9)$$

Also, it is easy to see from (3.8) that neither Jeffreys' prior π^J nor the reference prior π^R is a second order probability matching prior.

3.2.4 HPD Matching Prior

Due to the orthogonality of μ with $(\sigma_1, \dots, \sigma_k)$, from equation (33) of DiCiccio and Stern (1994) or equation (4.1) of Ghosh and Mukerjee (1995), a prior π is a HPD matching prior if and only if it satisfies

$$\frac{\partial^2}{\partial \mu^2} \left(I_{11}^{-1} \pi \right) - \frac{\partial}{\partial \mu} \left(I_{11}^{-2} L_{111} \pi \right) - \sum_{i=1}^k \frac{\partial}{\partial \sigma_i} \left[\left(\sigma_i^2 / 2n_i \right) I_{11}^{-1} \pi \right] = 0 \quad , \quad (3.10)$$

where $L_{111} = E\left[\frac{\partial^3 \log \mathcal{L}}{\partial \mu^3} \middle| \mu, \sigma_1, \dots, \sigma_k\right]$. Since $L_{111} = 0$ in this case, and the first term in the left hand side of (3.10) is 0, for any prior π of the form (3.6), one needs to solve (3.8) once again. Thus, π^S is a HPD matching prior, but π^J or π^R are not.

3.2.5 Matching via Inversion of Posterior Bartlett Corrected Likelihood Ratio Statistics

It follows from equations (2.3) and (2.4) of Yin and Ghosh (1997) that the differential equation leading to the characterization of such priors is the same for both unconditional and conditional Bartlett adjusted likelihood ratio statistic. Since $L_{1,1,1} = 0$ in this case, it follows from Theorem 1 of Yin and Ghosh (1997) that the second order probability matching prior derived in (3.9) also has the property that the difference between Bayesian and frequentist Bartlett corrections for the conditional likelihood ratio statistic tends to zero as $N \rightarrow \infty$. This property is not shared by either π^J or π^R .

3.3 Propriety of Posteriors

Since the priors π^R , π^S and π^J are all improper, the first thing is to establish the propriety of posteriors under these priors. This is what we accomplish in this section. Indeed, if the posterior is improper, it is pointless to talk about descriptive measures such as the mean, median, quantiles, as well as point estimates, credible intervals etc.

We begin with π^R . Then, the joint posterior of $\mu, \sigma_1, \dots, \sigma_k$ given $\mathbf{X}$ based on the reference prior π^R is

$$\pi^R(\mu, \sigma_1, \dots, \sigma_k | \mathbf{X}) \propto \prod_{j=1}^k \sigma_j^{-(n_j+1)} \exp \left[-\frac{1}{2} \sum_{j=1}^k \sigma_j^{-2} \left\{ n_j (\mu - \bar{X}_j)^2 + (n_j - 1) S_j^2 \right\} \right] . \quad (3.11)$$

Next, integrating with respect to $\sigma_1, \dots, \sigma_k$, the marginal posterior of μ given $\mathbf{X}$ based on the reference prior π^R is

$$\begin{aligned}
\pi^R(\mu|\mathbf{X}) &\propto \prod_{j=1}^k \left[n_j(\mu - \bar{X}_j)^2 + (n_j - 1)S_j^2 \right]^{-\frac{1}{2}n_j} \\
&\propto \prod_{j=1}^k \left[1 + \frac{n_j(\mu - \bar{X}_j)^2}{(n_j - 1)S_j^2} \right]^{-\frac{1}{2}n_j} .
\end{aligned} \tag{3.12}$$

This implies

$$\pi^R(\mu|\mathbf{X}) \leq C \left[1 + \frac{n_1(\mu - \bar{X}_1)^2}{(n_1 - 1)S_1^2} \right]^{-\frac{1}{2}n_1} , \tag{3.13}$$

where $C(> 0)$ depends on $(\bar{X}_1, \dots, \bar{X}_k, S_1^2, \dots, S_k^2)$, but not on μ . Now, the finiteness of $\int_{-\infty}^{\infty} \pi^R(\mu|\mathbf{X}) d\mu$ follows by recognizing the right hand side of (3.13) as a constant multiple of a Student's t -pdf.

Next, we prove the propriety of $\pi^S(\mu, \sigma_1, \dots, \sigma_k|\mathbf{X})$ based on the second order probability matching prior π^S . First, we write

$$\begin{aligned}
\pi^S(\mu, \sigma_1, \dots, \sigma_k|\mathbf{X}) &\propto \exp \left[-\frac{1}{2} \sum_{m=1}^k \sigma_m^{-2} \left\{ n_m(\mu - \bar{X}_m)^2 + (n_m - 1)S_m^2 \right\} \right] \\
&\times \prod_{j=1}^k \sigma_j^{-(n_j-1)} \left(\sum_{j=1}^k n_j \sigma_j^{-2} \right) \\
&= \exp \left[-\frac{1}{2} \sum_{m=1}^k \sigma_m^{-2} \left\{ n_m(\mu - \bar{X}_m)^2 + (n_m - 1)S_m^2 \right\} \right] \\
&\times \sum_{j=1}^k n_j \sigma_j^{-(n_j+1)} \left\{ \prod_{i \neq j}^k \sigma_i^{-(n_i-1)} \right\} .
\end{aligned} \tag{3.14}$$

Next, integrating with respect to $\sigma_1, \dots, \sigma_k$, the marginal posterior of μ given $\mathbf{X}$ based on the second order probability matching prior π^S is

$$\pi^S(\mu|\mathbf{X}) \propto \sum_{j=1}^k n_j \left\{ n_j(\mu - \bar{X}_j)^2 + (n_j - 1)S_j^2 \right\}^{-\frac{1}{2}n_j}$$

$$\begin{aligned}
& \times \prod_{i \neq j}^k \left\{ n_i (\mu - \bar{X}_i)^2 + (n_i - 1) S_i^2 \right\}^{-\frac{1}{2}(n_i - 2)} \\
& = \sum_{j=1}^k n_j \left\{ (n_j - 1) S_j^2 \right\}^{-\frac{1}{2}n_j} \prod_{i \neq j}^k \left\{ (n_i - 1) S_i^2 \right\}^{-\frac{1}{2}(n_i - 2)} \\
& \times \left[1 + \frac{n_j (\mu - \bar{X}_j)^2}{(n_j - 1) S_j^2} \right]^{-\frac{1}{2}n_j} \prod_{i \neq j}^k \left[1 + \frac{n_i (\mu - \bar{X}_i)^2}{(n_i - 1) S_i^2} \right]^{-\frac{1}{2}(n_i - 2)}. \quad (3.15)
\end{aligned}$$

Thus,

$$\pi^S(\mu | \mathbf{X}) \leq C_* \sum_{j=1}^k n_j \left\{ (n_j - 1) S_j^2 \right\}^{-\frac{1}{2}n_j} \prod_{i \neq j}^k \left\{ (n_i - 1) S_i^2 \right\}^{-\frac{1}{2}(n_i - 2)} \left[1 + \frac{n_j (\mu - \bar{X}_j)^2}{(n_j - 1) S_j^2} \right]^{-\frac{1}{2}n_j}, \quad (3.16)$$

where $C_*(> 0)$ depends on the data, but not on μ . Once again, from (3.16), the finiteness of $\int_{-\infty}^{\infty} \pi^S(\mu | \mathbf{X}) d\mu$ follows from the finiteness of t -integrals.

Finally, the propriety of $\pi^J(\mu, \sigma_1, \dots, \sigma_k | \mathbf{X})$ follows by arguments similar to those used in proving the finiteness of $\int_0^\infty \dots \int_0^\infty \int_{-\infty}^\infty \pi^R(\mu, \sigma_1, \dots, \sigma_k | \mathbf{X}) d\mu d\sigma_1 \dots d\sigma_k$.

3.4 Simulation Study

We compare in this section the two priors π^R and π^S by calculating the frequentist probabilities $P[\mu \leq \mu^{1-\alpha}(\pi^R, \mathbf{X}) | \mu, \sigma_1, \sigma_2]$ and $P[\mu \leq \mu^{1-\alpha}(\pi^S, \mathbf{X}) | \mu, \sigma_1, \sigma_2]$. We write $F^{\pi^R}(z) = P^{\pi^R}(\mu \leq z | \mathbf{X})$ and $F^{\pi^S}(z) = P^{\pi^S}(\mu \leq z | \mathbf{X})$. Thus, $F^{\pi^R}(\mu^{1-\alpha}(\pi^R; \mathbf{X})) = 1 - \alpha = F^{\pi^S}(\mu^{1-\alpha}(\pi^S; \mathbf{X}))$.

Let $P(\alpha; \pi^R, \mu, \sigma_1, \sigma_2) = P(\mu \leq \mu^{1-\alpha}(\pi^R, \mathbf{X}) | \mu, \sigma_1, \sigma_2)$ and $P(\alpha; \pi^S, \mu, \sigma_1, \sigma_2) = P(\mu \leq \mu^{1-\alpha}(\pi^S, \mathbf{X}) | \mu, \sigma_1, \sigma_2)$. If the quantities $P(\alpha; \pi^R, \mu, \sigma_1, \sigma_2)$ and $P(\alpha; \pi^S, \mu, \sigma_1, \sigma_2)$ are close to $1 - \alpha$ even if the sample sizes are small, then there is evidence that these priors perform well with respect to the probability matching criterion.

We generate a variety of samples of sizes n_1 and n_2 from $\mathcal{N}(\mu, \sigma_1^2)$ and $\mathcal{N}(\mu, \sigma_2^2)$ distributions. Throughout, we take $\mu = 0$ and $\sigma_2^2 = 1$, but take different values of σ_1^2 . The results in Tables 1 and 2 are reported for $\sigma_1^2 = 2$ and $\sigma_1^2 = 4$.

$P(\alpha; \pi^R, \mu, \sigma_1, \sigma_2)$ and $P(\alpha; \pi^S, \mu, \sigma_1, \sigma_2)$ are estimated in the following way. For each pair (n_1, n_2) , 10,000 random samples $(x_{11}, \dots, x_{1n_1})$ and $(x_{21}, \dots, x_{2n_2})$ are generated using FORTRAN77 software. Next $F^{\pi^R}(\mu)$ and $F^{\pi^S}(\mu)$ are computed for each of the 10,000 sets of random samples. Then $P(\alpha; \pi^R, \mu, \sigma_1, \sigma_2)$ and $P(\alpha; \pi^S, \mu, \sigma_1, \sigma_2)$ are estimated by the proportions of times $F^{\pi^R}(\mu)$ and $F^{\pi^S}(\mu)$ are less than or equal to $1 - \alpha$.

It is clear from the following Tables that π^S matches the target much more accurately than π^R . This is intuitively expected since the former is a second order matching prior, but the latter is not.

3.5 Comparison with Other Frequentist Methods

We introduce in this section some existing frequentist methods for construction of confidence intervals for μ and compare the performance of the second order probability matching prior against these by analyzing two real datasets. These frequentist methods are all discussed in Yu, Sun and Sinha (1999).

Recall that $\bar{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$ and $(n_i - 1)S_i^2 = \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$, $i = 1, \dots, k$. Since

$$T_i = \frac{\sqrt{n_i}(\bar{X}_i - \mu)}{S_i} \sim t_{n_i-1} \quad (3.17)$$

or, equivalently,

$$F_i = \frac{n_i(\bar{X}_i - \mu)^2}{S_i^2} \sim F_{1, n_i-1} \quad (3.18)$$

are standard test statistics for testing hypotheses about μ based on the i th sample, suitable linear combinations of $|T_i|$'s or F_i 's or other functions thereof have been suggested often as *pivots* to construct exact confidence intervals for μ . This is precisely

Table 3.1: Frequentist coverage probabilities of 0.05 (0.95) posterior quantiles of μ when $\mu = 0.0$ and $\sigma_2^2 = 1.0$ ($n_1 > n_2$)

n_1	n_2	π	$\sigma_1^2 = 2.0$		$\sigma_1^2 = 4.0$	
			0.05	0.95	0.05	0.95
3	2	π^R	0.097	0.908	0.095	0.912
		π^S	0.050	0.951	0.050	0.950
7	5	π^R	0.070	0.934	0.068	0.936
		π^S	0.054	0.949	0.054	0.949
15	10	π^R	0.057	0.945	0.057	0.944
		π^S	0.051	0.951	0.051	0.949
20	15	π^R	0.054	0.944	0.054	0.945
		π^S	0.049	0.949	0.050	0.950
30	20	π^R	0.056	0.945	0.054	0.945
		π^S	0.053	0.948	0.051	0.949

Table 3.2: Frequentist coverage probabilities of 0.05 (0.95) posterior quantiles of μ when $\mu = 0.0$ and $\sigma_2^2 = 1.0$ ($n_1 < n_2$)

n_1	n_2	π	$\sigma_1^2 = 2.0$		$\sigma_1^2 = 4.0$	
			0.05	0.95	0.05	0.95
2	3	π^R	0.095	0.916	0.086	0.921
		π^S	0.055	0.949	0.055	0.949
5	7	π^R	0.065	0.936	0.061	0.941
		π^S	0.050	0.951	0.049	0.951
10	15	π^R	0.058	0.945	0.056	0.946
		π^S	0.053	0.950	0.051	0.950
15	20	π^R	0.055	0.946	0.054	0.945
		π^S	0.052	0.948	0.050	0.949
20	30	π^R	0.055	0.946	0.053	0.946
		π^S	0.052	0.948	0.050	0.948

what is accomplished in Cohen and Sackrowitz (1984), and in Jordan and Krishnamoorthy (1996). Their work is summarized below.

3.5.1 Confidence Interval for μ Based on T_i 's

Cohen and Sackrowitz (1984) suggested use of $M_t = \max_{1 \leq i \leq k} \{|T_i|\}$ as a test statistic for testing hypotheses about μ . We can use M_t to construct a confidence interval for μ once the cutoff point of the distribution of M_t is known. This cutoff point under the null hypothesis is denoted by $c_{\alpha/2}$. Thus, if $c_{\alpha/2}$ satisfies the condition

$$\begin{aligned} 1 - \alpha &= P[M_t \leq c_{\alpha/2}] \\ &= \prod_{i=1}^k P[|T_i| \leq c_{\alpha/2}] , \end{aligned}$$

an exact confidence interval for μ with confidence level $1 - \alpha$ is given by

$$\left[\max_{1 \leq i \leq k} \left\{ \bar{X}_i - \frac{c_{\alpha/2} S_i}{\sqrt{n_i}} \right\} , \min_{1 \leq i \leq k} \left\{ \bar{X}_i + \frac{c_{\alpha/2} S_i}{\sqrt{n_i}} \right\} \right] \quad (3.19)$$

Determination of the cutoff point $c_{\alpha/2}$ is not easy in applications, and simulation may be necessary. An alternative approach is to use the confidence interval

$$\left[\max_{1 \leq i \leq k} \left\{ \bar{X}_i - \frac{c_{\alpha/2}^{(i)} S_i}{\sqrt{n_i}} \right\} , \min_{1 \leq i \leq k} \left\{ \bar{X}_i + \frac{c_{\alpha/2}^{(i)} S_i}{\sqrt{n_i}} \right\} \right] \quad (3.20)$$

where $c_{\alpha/2}^{(i)}$ satisfies $P[|T_i| \leq c_{\alpha/2}^{(i)}] = (1 - \alpha)^{1/k}$. This latter interval clearly also has an exact coverage probability $1 - \alpha$.

Fairweather (1972) suggested using a weighted linear combinations of T_i 's, namely,

$$W_t = \sum_{i=1}^k u_i T_i, \quad u_i = \frac{[\text{Var}(T_i)]^{-1}}{\sum_{j=1}^k [\text{Var}(T_j)]^{-1}} \quad (3.21)$$

which is also a *pivot*. If $b_{\alpha/2}$ denotes the cutoff point of the distribution W_t , satisfying the equation

$$1 - \alpha = P[|W_t| \leq b_{\alpha/2}] \quad , \quad (3.22)$$

then the confidence interval for μ is obtained as

$$\left[\frac{\sum_{i=1}^k \sqrt{n_i} u_i \bar{X}_i / S_i}{\sum_{i=1}^k \sqrt{n_i} u_i / S_i} - \frac{b_{\alpha/2}}{\sum_{i=1}^k \sqrt{n_i} u_i / S_i} \quad , \quad \frac{\sum_{i=1}^k \sqrt{n_i} u_i \bar{X}_i / S_i}{\sum_{i=1}^k \sqrt{n_i} u_i / S_i} + \frac{b_{\alpha/2}}{\sum_{i=1}^k \sqrt{n_i} u_i / S_i} \right] \quad (3.23)$$

It may be noted that u_i 's can be easily derived using the fact that

$$var(T_\nu) = \frac{\nu}{\nu - 2} \quad , \quad \nu > 2 \quad .$$

3.5.2 Confidence Interval for μ Based on F_i 's

Jordan and Krishnamoorthy (1996) suggested using linear combinations of the F_i 's such as $W_f = \sum_{i=1}^k w_i F_i$ for positive weight w_i 's, which is again the *pivot*. Hence, if we can compute a_α such that

$$P[W_f \leq a_\alpha] = 1 - \alpha \quad ,$$

then, after simplification, an exact confidence interval for μ with confidence level $1 - \alpha$ is given by

$$\left[\sum_{i=1}^k p_i \bar{X}_i - \Delta_i \quad , \quad \sum_{i=1}^k p_i \bar{X}_i + \Delta_i \right] \quad (3.24)$$

where

$$p_i = \frac{w_i n_i / S_i^2}{\sum_{j=1}^k w_j n_j / S_j^2}$$

and

$$\Delta_i^2 = \frac{a_\alpha}{\sum_{i=1}^k w_i n_i / S_i^2} - \sum_{i=1}^k p_i \bar{X}_i^2 + \left\{ \sum_{i=1}^k p_i \bar{X}_i \right\}^2$$

Jordan and Krishnamoorthy (1996) used w_i as inversely proportional to $\text{var}(F_i) = 2m_i^2(m_i - 1)/[(m_i - 2)^2(m_i - 4)]$ where $m_i = n_i - 1$, resulting in w_i as

$$w_i = \frac{[(m_i - 2)^2(m_i - 4)]/[m_i^2(m_i - 1)]}{\sum_{j=1}^k [(m_j - 2)^2(m_j - 4)]/[m_j^2(m_j - 1)]}$$

Of course, it is assumed that $n_i > 5$ for all the k populations.

3.5.3 Confidence Regions for μ Based on P_i 's

Since F_i defined in (3.18) can be used for testing hypotheses about μ , one defines the i th P value, P_i as

$$P_i = \int_{F_i}^{\infty} f_i(x) dx$$

where $f_i(x)$ denotes the pdf of the F distribution with 1 and $(n_i - 1)$ degrees of freedom. Recalling the fact that $P_1, \dots, P_k$ are *iid* uniformly distributed random variables, one can combine them using several well known methods (Hedges and Olkin, 1985). In particular, we discuss below methods such as Tippett's (Tippett, 1931), Fisher's (Fisher, 1932), inversion normal (Stouffer, Suchman, DeVinney, Star and Williams, 1949), and logit (George, 1977) for data analysis, and compare them with the Bayesian procedure.

Tippett's Method

If $P_{[1]}$ is the minimum of $P_1, P_2, \dots, P_k$, then Tippett's method rejects the hypothesis about μ if $P_{[1]} < c_1 = 1 - (1 - \alpha)^{1/k}$. By inverting this rejection region, we have

a confidence interval for μ with confidence coefficient $1 - \alpha$, given by

$$\begin{aligned}
 \text{C.I} &= \{\mu : P_{[1]} \geq c_1\} \\
 &= \{\mu : P_i \geq c_1, i = 1, \dots, k\} \\
 &= \{\mu : \int_{n_i(\bar{X}_i - \mu)^2 / S_i^2}^{\infty} f_i(x) dx \geq 1 - (1 - \alpha)^{1/k}, i = 1, \dots, k\}. \quad (3.25)
 \end{aligned}$$

It may be noted that (3.25) exactly coincides with (3.20).

Fisher's Method

Since Fisher's method rejects hypotheses about μ when $-2 \sum_{i=1}^k \log P_i > \chi_{2k, \alpha}^2$, the confidence interval for μ obtained by inverting the acceptance region of this test is given by

$$\begin{aligned}
 \text{C.I} &= \{\mu : -2 \sum_{i=1}^k \log P_i \leq \chi_{2k, \alpha}^2\} \\
 &= \{\mu : \prod_{i=1}^k P_i \geq e^{-\chi_{2k, \alpha}^2 / 2}\} \\
 &= \{\mu : \prod_{i=1}^k \int_{n_i(\bar{X}_i - \mu)^2 / S_i^2}^{\infty} f_i(x) dx \geq e^{-\chi_{2k, \alpha}^2 / 2}\}. \quad (3.26)
 \end{aligned}$$

Inversion Normal Method

Since this method rejects hypotheses about μ when $\frac{\sum_{i=1}^k \Phi^{-1}(P_i)}{\sqrt{k}} < -z_\alpha$ at level α , the $1 - \alpha$ level confidence interval for μ obtained by inverting this acceptance region is given by

$$\text{C.I} = \left\{ \mu : \frac{\sum_{i=1}^k \Phi^{-1}(P_i)}{\sqrt{k}} < -z_\alpha \right\}, \quad (3.27)$$

where $\Phi(\cdot)$ is a cdf of standard normal distribution.

Logit Method

This method rejects H_0 if $\sum_{i=1}^k \log \frac{P_i}{1-P_i} < c$ where c is a predetermined constant. It was shown by George and Mudholkar (1977) that the distribution of

$$G^* = \left[-\sum_{i=1}^k \log \frac{P_i}{1-P_i} \right] \left[\frac{3}{k\pi^2} \right]^{1/2}$$

can be closely approximated by a standard normal distribution. Therefore a $(1 - \alpha)$ level confidence interval for μ can be obtained from

$$\text{C.I} = \{\mu : G^* \leq z_\alpha\} \quad (3.28)$$

3.5.4 Data Analysis

In order to see and compare the performance of the second order probability matching prior derived in Section 3.2 with the frequentist methods described in the previous section, we analyze two sets of real-life data.

Example 1 In an example given by Snedecor (1950) the data from four experiments are used to estimate the percentage of albumin in plasma protein of normal human subjects. This dataset is also reported in Meier (1953) and is analyzed in Jordan and Krishnamoorthy (1996). We would like to combine the results of four experiments in order to construct a confidence interval for the common mean μ . The data appear in Table 3.3.

We have applied all the methods described in Section 5, and computed the two-sided confidence intervals for μ with $\alpha = 0.05$. Also, we have computed the two-sided confidence intervals and HPD regions based on the second order probability matching priors to compare its performance at the same significance level. Figure 3.1 illustrates the marginal posterior pdf of the common mean μ corresponding to the

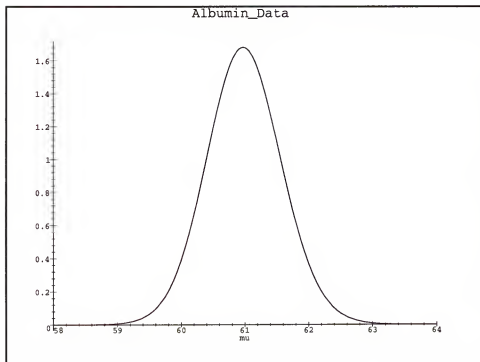
Table 3.3: Percentage of albumin in plasma protein

Experiment	n_i	Mean	Variance
A	12	62.3	12.986
B	15	60.3	7.840
C	7	59.5	33.433
D	16	61.5	18.513

second order probability matching prior. Table 3.4 gives the resulting equal two-sided confidence intervals and HPD regions. It turns out that the confidence intervals based on Bayesian procedures are comparable to the best frequentist procedure in the sense of having the smallest observed length. The best frequentist procedure here is the one due to Fairweather (1972) where the confidence interval has length 2.30. The Bayes confidence interval has length 2.31.

Table 3.4: Interval Estimates for μ

Methods	Critical Values	Weights	Interval
C & S (3.19)	$c=0.3043$		[59.14, 62.50]
T (3.20)	$c_i=2.9702, 2.8543,$ $3.5055, 2.8272$		[59.20, 62.36]
F (3.23)	$b \doteq 1.102$	$u_i=0.2550, 0.2671,$ $0.2708, 0.2701$	[59.89, 62.19]
J & K (3.24)	$a \doteq 3.191$	$p_i=0.2100, 0.5245,$ $0.0181, 0.2474$	[59.56, 62.44]
Fisher (3.26)			[59.47, 62.56]
Inverse Normal (3.27)			[59.27, 62.66]
Logit (3.28)			[59.29, 62.61]
Prob.Matching Prior			
Equal Two-sided			[59.85, 62.16]
HPD region			[59.86, 62.17]

Figure 3.1: Posterior pdf of μ

Example 2 This example appears in Eberhardt, Reeve and Spiegelman (1989) and deals with the problem of estimation of mean *Selenium* in non-fat milk powder by combining the results of four methods. The data appear in Table 3.5.

Table 3.5: Selenium in non-fat milk powder

Methods	n_i	Mean	Variance
Atomic absorption spectrometry	8	105.00	85.711
Neutron activation			
1) Instrumental	12	109.75	20.748
2) Radiochemical	14	109.50	2.729
Isotope dilution mass spectrometry	8	113.25	33.640

Here again we have applied all the methods described in Section 5, and have computed the two-sided confidence intervals for μ with $\alpha = 0.05$. Also, we have computed

the two-sided confidence intervals and HPD regions based on the second-order probability matching priors to compare the performance at the same significance level. The marginal posterior pdf of the common mean μ corresponding to the second order probability matching prior appears in Figure 3.2. Table 3.6 gives the resulting equal two-sided confidence intervals and HPD regions. As a result, the confidence intervals based on Bayesian procedures seems to be near equivalent to the best frequentist procedure based on inverse normal, but it is superior to other frequentist procedures. Also, unlike the previous example the Fairweather interval does not have the shortest length here.

Table 3.6: Interval Estimates for μ

Methods	Critical Values	Weights	Interval
C & S (3.19)	$c=3.1283$		[108.12, 110.88]
T (3.20)	$c_t=3.321, 2.970,$ 2.886, 3.321		[108.23, 110.77]
F (3.23)	$b \doteq 1.118$	$u_i=0.2309, 0.2645,$ 0.2736, 0.2309	[108.59, 110.81]
J & K (3.24)	$a \doteq 3.341$	$p_i=0.0068, 0.0777,$ 0.8908, 0.0247	[108.52, 110.68]
Fisher (3.26)			[108.59, 110.61]
Inverse Normal (3.27)			[108.78, 110.47]
Logit (3.28)			[108.72, 110.51]
Prob.Matching Prior			
Equal Two-sided			[108.69, 110.48]
HPD region			[108.71, 110.49]

3.6 Concluding Remarks

The chapter presents yet another example where a unified Bayesian procedure can lead to good frequentist properties in situations where there is no clear-cut choice

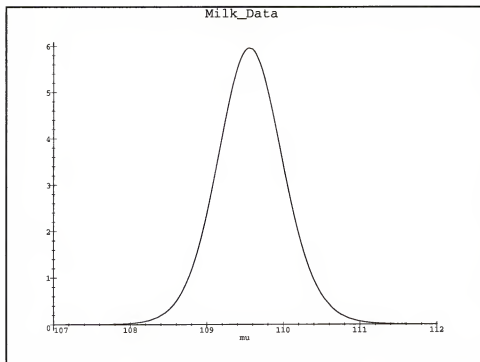


Figure 3.2: Posterior pdf of μ

between frequentist procedures. Yu, Sun and Sinha (1999) have recommended the use of Fisher's method for finding confidence intervals for the common mean based primarily on pragmatic considerations. In both the examples considered, the proposed Bayesian credible interval has smaller lengths than the interval based on Fisher's method. Thus, the Bayesian procedure should be given every serious consideration even for frequentist analysis.

CHAPTER 4
BAYESIAN INFERENCE FOR THE RATIOS OF REGRESSION
COEFFICIENTS IN A LINEAR MODEL

4.1 Introduction

There is a large class of important statistical problems which can be categorized under the general heading of inference about the ratio of two regression coefficients in a general linear model. Included in this class are (i) the calibration problem, (ii) the Fieller-Creasy problem, (iii) parallel-line bioassay, and (iv) slope-ratio bioassay.

The objective of this chapter is to present a unified default Bayesian analysis for the general problem involving ratios of regression coefficients. To this end, once again an orthogonal transformation of the parameter vector is proposed. This orthogonal transformation becomes a convenient starting point in the development of a variety of default Bayesian priors.

The general regression model is given by

$$Y_i = \sum_{j=1}^r \beta_j x_{ij} + \epsilon_i \quad , \quad (i = 1, \dots, n) \quad (4.1)$$

where the parameter of interest is $\rho = \beta_1/\beta_2$. In matrix notation, we write the model as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad , \quad (4.2)$$

where

$$\mathbf{Y} = (Y_1, \dots, Y_n)^T \quad ,$$

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1r} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nr} \end{pmatrix}$$

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)^T, \text{ and}$$

$$\mathbf{e} = (e_1, \dots, e_n)^T.$$

We also assume that $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_n)$.

In Section 4.2, we introduce some special examples of important statistical problems which all arise as special cases of the general problem discussed above. This observation is not new. For instance, these models have all been previously recognized as special cases of the general model by Buonaccorsi and Gatsonis (1988).

The new feature of this chapter is the introduction of a nontrivial orthogonal transformation of the original parameter vector $(\boldsymbol{\beta}, \sigma)$ in Section 4.3 which facilitates systematic development of default Bayesian priors, such as the different reference priors and probability matching priors. In this process, some of the priors proposed for example in Liseo (1993), Yin and Ghosh (1997), Hunter and Lamboy (1981), Ghosh, Carlin and Srivastava(1995), Mendoza (1996) and many others are formed as special cases, and more important, some new priors are developed.

The reference priors and probability matching priors are developed in Section 4.4. Section 4.5 proves the propriety of posteriors under these default priors with very mild assumptions on the sample sizes. Two real examples of special cases, parallel-line bioassay and slope-ratio bioassay, are discussed in Section 4.6. Section 4.7 provides mathematical proofs of theorems.

4.2 Some Special Examples

4.2.1 The Linear Calibration Problem

Statistical calibration problems arise in the presence of two measurement methods and involves making inferences about fixed but unknown explanatory variables corresponding to a response variable. Since the appearance of Krutchkoff's (1967) controversial paper, a large number of papers have appeared on the topic. Univariate Bayesian calibration has been discussed by Hoadley (1970) and Hunter and Lamboy (1981). Hoadley (1970) showed that Krutchkoff's (1967) inverse regression estimator can be interpreted in a Bayesian way. Hunter and Lamboy (1981), on the other hand, tended to provide Bayesian support of the classical estimator. The prior of Hunter and Lamboy (1981) was criticized both by Hill (1981) and Lawless (1981). In particular, Hill (1981) criticized the authors for introducing a prior "with no motivation and no attempt to understand either its implementations or its relationship with other prior distributions". We will see, however, that the Hunter-Lamboy prior is indeed a second order probability matching prior.

Ghosh, Carlin and Srivastava (1995) discussed a Bayesian approach to univariate calibration with the reference priors and the first order probability matching priors. They show that Jeffreys noninformative prior as well as the one used by Hunter and Lamboy (1981) are first order probability matching priors.

In a univariate linear calibration problem, the calibration experiment and the prediction experiment can be represented as

$$\begin{aligned} Z_{1i} &= \alpha + \beta x_i + \epsilon_{1i} \quad (i = 1, \dots, m) \\ Z_{2j} &= \alpha + \beta \rho + \epsilon_{2j} \quad (j = 1, \dots, k) \quad , \end{aligned} \tag{4.3}$$

respectively, where ϵ_{1i} and ϵ_{2j} are i.i.d $\mathcal{N}(0, \sigma^2)$. This is a special case of the model given in (4.1) with $n = m + k$, $r = 3$, $Y_i = Z_{1i} (i = 1, \dots, m)$, $Y_{m+j} = Z_{2j} (j = 1, \dots, k)$. Also, $\beta_1 = \beta\rho$, $\beta_2 = \beta$, $\beta_3 = \alpha$ and the design matrix $\mathbf{X}$ is given by

$$\mathbf{X}^T = \begin{bmatrix} 0 & \cdots & 0 & 1 & \cdots & 1 \\ x_1 & \cdots & x_m & 0 & \cdots & 0 \\ 1 & \cdots & 1 & 1 & \cdots & 1 \end{bmatrix} . \quad (4.4)$$

4.2.2 The Fieller-Creasy Problem

The celebrated Fieller-Creasy problem (Fieller, 1954; Creasy, 1954) involves inference about the ratio of two normal means. This problem has posed a constant challenge to frequentist and likelihood based inference. Fieller's method of providing a confidence set for this ratio based on a pivot can lead to two disjoint unbounded sets or even the whole real line. Also, as pointed out by Gleser and Hwang (1987) (see also Berger, Liseo and Wolpert, 1996) based on any sample of arbitrary but fixed size n , a confidence interval of finite expected length for this ratio has coverage probability (taking the inf. over all points in the parameter space) of 0.

Bayesian analysis for this problem based on noninformative priors began with Kappenman, Geisser and Antle (1970), and was addressed subsequently in Bernardo (1977), Sendra (1982), Mendoza (1988, 1996), Stephens and Smith (1992), Liseo (1993), Phillipe and Robert (1998), and Berger, Liseo and Wolpert (1996).

Liseo (1993) compared profile likelihood and its modification with a Bayesian analysis based on reference priors, when the variance is known. Mendoza (1996) investigated frequentist coverage probabilities of HPD intervals for reference priors through simulations with known variance. In general, integrated likelihood approaches with noninformative priors are advocated by Berger, Liseo and Wolpert (1996) for tackling a class of important but troublesome problems, including the Fieller-Creasy problem.

The assumed model for the Fieller-Creasy problem is given by

$$\begin{aligned} Z_{1i} &= \beta + \epsilon_{1i} \quad (i = 1, \dots, m) \\ Z_{2j} &= \beta\rho + \epsilon_{2j} \quad (j = 1, \dots, k) \quad , \end{aligned} \quad (4.5)$$

where ϵ_{1i} and ϵ_{2j} are i.i.d $\mathcal{N}(0, \sigma^2)$. This is once again a special case of the model given in (4.1) with $n = m + k$, $r = 2$, $Y_i = Z_{1i} (i = 1, \dots, m)$, $Y_{m+j} = Z_{2j} (j = 1, \dots, k)$, $\beta_1 = \beta\rho$, $\beta_2 = \beta$. The design matrix $\mathbf{X}$ is given by

$$\mathbf{X}^T = \begin{bmatrix} 0 & \cdots & 0 & 1 & \cdots & 1 \\ 1 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix} \quad . \quad (4.6)$$

4.2.3 Parallel-Line Bioassay Problem

Consider a test drug T being compared to a standard drug S . T and S are said to be similar if there is a constant $\mu > 0$ such that the distribution of the response to dose w_2 of the test drug is equivalent to that of dose $w_1 = \mu w_2$ of the standard drug. When the drugs are similar, μ is called the relative potency of T to S .

In many dose-response experiments the quantitative response Z is linearly related to the log dose, $x = \log w$, and we can assume the model

$$E[Z|x] = \alpha + \beta x \quad (4.7)$$

where α and β are unknown real-valued parameters. If S and T are similar, then the responses Z_1 and Z_2 to S and T , respectively, are related by

$$\begin{aligned} E[Z_1|x] &= \alpha + \beta x \\ E[Z_2|x] &= \alpha + \beta(x + \rho), \end{aligned} \quad (4.8)$$

for some α, β and ρ , where $\rho = \log \mu$. This follows from the assumption that dose w of T is equivalent to dose μw of S . The estimation of μ (or equivalently ρ) based on this model is called a parallel-line bioassay.

Given the intercepts $\alpha_1 = \alpha$ and $\alpha_2 = \alpha + \beta \log \mu$ and the common slope β , we may observe the horizontal distance between the two regression lines in (4.8) is $\rho = \log \mu = (\alpha_2 - \alpha_1)/\beta$.

Consider an experiment where p doses $(x_{11}, \dots, x_{1p})$ of a standard drug S is assayed m times and q doses $(x_{21}, \dots, x_{2q})$ of a test drug T is assayed u times so that a set $\{Z_{1ik}, i = 1, \dots, p; k = 1, \dots, m; Z_{2jk}, j = 1, \dots, q; k = 1, \dots, u\}$ of $pm + qu$ observations are obtained. The assumed model here is

$$\begin{aligned} Z_{1ik} &= \alpha + \beta x_{1i} + \epsilon_{1ik}, & (k = 1, \dots, m; i = 1, \dots, p) \\ Z_{2jk} &= \alpha + \beta(x_{2j} + \rho) + \epsilon_{2jk}, & (k = 1, \dots, u; j = 1, \dots, q) \end{aligned} \quad (4.9)$$

where ϵ_{1ik} and ϵ_{2jk} are i.i.d $\mathcal{N}(0, \sigma^2)$. Once again, ρ is the parameter of interest.

In order to recognize this model as a special case of (4.1), we first represent the Z_{1ik} 's and Z_{2jk} 's as the $\mathbf{Y}$ vector as in the previous example. Also, as before $n = pm + qu$, $r = 3$, $\beta_1 = \beta\rho$, $\beta_2 = \beta$ and $\beta_3 = \alpha$. The design matrix $\mathbf{X}$ in this case is given by

$$\mathbf{X}^T = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 1 & \cdots & 1 & \cdots & 1 & \cdots & 1 \\ x_{11} & \cdots & x_{11} & \cdots & x_{1p} & \cdots & x_{1p} & x_{21} & \cdots & x_{21} & \cdots & x_{2q} & \cdots & x_{2q} \\ 1 & \cdots & 1 & \cdots & 1 & \cdots & 1 & 1 & \cdots & 1 & \cdots & 1 & \cdots & 1 \end{bmatrix}. \quad (4.10)$$

4.2.4 Slope-Ratio Bioassay Problem

Estimation of slope-ratio is a common problem in biological assays. Consider an experiment where p doses $(x_{11}, \dots, x_{1p})$ of a standard drug S is assayed m times

and q doses $(x_{21}, \dots, x_{2q})$ of a test drug T is assayed u times so that a set $\{Z_{1ik}, i = 1, \dots, p; k = 1, \dots, m; Z_{2jk}, j = 1, \dots, q; k = 1, \dots, u\}$ of $pm + qu$ observations are obtained.

The classical approach to the statistical analysis of these ratio-type problems has the common drawbacks as given in Fieller (1954), Creasy (1954). On the other hand, likelihood based approach will also encounter difficulties in producing interval estimators. Mendoza (1990) discussed Bayesian analysis for the slope-ratio problem. Under the normality assumption, he derived the two-group reference prior and compared the corresponding credible intervals with the confidence intervals via Fieller's theorem.

The assumed model for slope-ratio bioassay is

$$\begin{aligned} Z_{1ik} &= \alpha + \beta x_{1i} + \epsilon_{1ik} \quad (k = 1, \dots, m; i = 1, \dots, p) \\ Z_{2jk} &= \alpha + \beta \rho x_{2j} + \epsilon_{2jk} \quad (k = 1, \dots, u; j = 1, \dots, q) \end{aligned} \quad (4.11)$$

where ϵ_{1ik} and ϵ_{2jk} are i.i.d $\mathcal{N}(0, \sigma^2)$. To see how this problem arises as a special case of the general regression model given in (4.1), let

$$\begin{array}{ccccccc} Y_1 = Z_{111}, & Y_2 = Z_{112}, & \dots, & Y_m = Z_{11m} \\ Y_{m+1} = Z_{121}, & Y_{m+2} = Z_{122}, & \dots, & Y_{2m} = Z_{12m} \\ \vdots & \vdots & \vdots & \\ Y_{(p-1)m+1} = Z_{1p1}, & Y_{(p-1)m+2} = Z_{1p2}, & \dots, & Y_{pm} = Z_{1pm} \\ Y_{pm+1} = Z_{211}, & Y_{pm+2} = Z_{212}, & \dots, & Y_{pm+u} = Z_{21u} \\ \vdots & \vdots & \vdots & \\ Y_{pm+(q-1)u+1} = Z_{2q1}, & Y_{pm+(q-1)u+2} = Z_{2q2}, & \dots, & Y_{pm+qu} = Z_{2qu} \end{array} .$$

We can then identify this model as a special case of the general regression model with $n = pm + qu$, $r = 3$ and $\beta_1 = \beta \rho$, $\beta_2 = \beta$, $\beta_3 = \alpha$. Also, the design matrix $\mathbf{X}$ is

given by

$$\mathbf{X}^T = \begin{bmatrix} 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & x_{21} \cdots x_{21} & \cdots & x_{2q} \cdots x_{2q} \\ x_{11} \cdots x_{11} & \cdots & x_{1p} \cdots x_{1p} & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 1 & \cdots & 1 & \cdots & 1 & \cdots & 1 & 1 & \cdots & 1 \end{bmatrix} . \quad (4.12)$$

4.3 The Orthogonal Transformation

This section introduces a transformation of the parameter vector $(\boldsymbol{\beta}, \sigma)$ which results in the orthogonality (Cox and Reid, 1987) of ρ with the remaining parameters.

We begin with the Fisher information matrix

$$\mathbf{I}(\boldsymbol{\beta}, \sigma) = \begin{bmatrix} ((s_{jl})) & \mathbf{0}^T \\ \mathbf{0} & 2n \end{bmatrix} ,$$

where $s_{jl} = \sum_{i=1}^n x_{ij}x_{il}$ ($j, l = 1, \dots, r$). Consider the transformation

$$\begin{aligned} \beta_1 &= \theta_1 \theta_2 h(\theta_1) \\ \beta_2 &= \theta_2 h(\theta_1) \\ \beta_j &= \theta_j - \theta_2 g_j(\theta_1) , \quad j = 3, \dots, r \\ \sigma &= \theta_{r+1} . \end{aligned}$$

Then the Jacobian matrix is given by

$$\mathbf{J} = \begin{bmatrix} \theta_2 \{ \theta_1 h'(\theta_1) + h(\theta_1) \} & \theta_2 h'(\theta_1) & -\theta_2 g'_3(\theta_1) & \cdots & -\theta_2 g'_r(\theta_1) & 0 \\ \theta_1 h(\theta_1) & h(\theta_1) & -g_3(\theta_1) & \cdots & -g_r(\theta_1) & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix} .$$

Theorem 4.1 Let

$$g_j(\theta_1) = h(\theta_1)(a_{j1}\theta_1 + a_{j2}) , \quad j = 3, \dots, r$$

where

$$\begin{bmatrix} a_{31} & a_{32} \\ \vdots & \vdots \\ a_{r1} & a_{r2} \end{bmatrix} = \begin{bmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{3r} & \cdots & s_{rr} \end{bmatrix}^{-1} \begin{bmatrix} s_{13} & s_{23} \\ \vdots & \vdots \\ s_{1r} & s_{2r} \end{bmatrix} .$$

Then parametric orthogonality holds by choosing

$$h(\theta_1) = [c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22}]^{-\frac{1}{2}} \quad (4.13)$$

where

$$\begin{aligned} c_{11} &= s_{11} - \sum_{j=3}^r s_{1j}a_{j1} \\ c_{12} &= s_{12} - \sum_{j=3}^r s_{1j}a_{j2} \\ c_{22} &= s_{22} - \sum_{j=3}^r s_{2j}a_{j2} . \end{aligned}$$

The proof of this theorem is deferred to the appendix.

The resulting Fisher information matrix is then given by

$$\mathbf{I}(\theta) = \frac{1}{\theta_{r+1}^2} \begin{bmatrix} \frac{\theta_1^2(c_{11}c_{22}-c_{12}^2)}{(c_{11}\theta_1^2+2c_{12}\theta_1+c_{22})^2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & s_{33} & \cdots & s_{3r} & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & s_{r3} & \cdots & s_{rr} & 0 \\ 0 & 0 & 0 & \cdots & 0 & 2n \end{bmatrix} \quad (4.14)$$

4.4 Development of Noninformative Priors

From (4.2) the likelihood function is given by

$$L(\theta) \propto \theta_{r+1}^{-n} \exp \left[-\frac{1}{2\theta_{r+1}^2} \sum_{i=1}^n \left\{ y_i - \theta_2 h(\theta_1) [\theta_1 x_{i1} + x_{i2} - \sum_{j=3}^r x_{ij} (a_{j1}\theta_1 + a_{j2})] - \sum_{j=3}^r x_{ij} \theta_j \right\}^2 \right]. \quad (4.15)$$

We now consider several noninformative priors similar to the ones given in the previous chapters.

4.4.1 Jeffreys' Prior

Jeffreys' prior π^J is proportional to the positive square root of the determinant of the $\mathbf{I}(\theta)$. Thus,

$$\pi^J(\theta) \propto \frac{|\theta_2|}{\theta_{r+1}^{r+1}} \left(c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22} \right)^{-1}. \quad (4.16)$$

4.4.2 Reference Priors

First we consider the two-group reference prior. The parameter space is partitioned into $\{\theta_1\}, \{\theta_2, \dots, \theta_{r+1}\}$, θ_1 being the parameter of interest. Let $\theta_{(1)} = \theta_1$ and $\theta_{(2)} = (\theta_2, \dots, \theta_{r+1})$. Then the Fisher information matrix can be expressed as

$$\mathbf{I}(\theta) = \text{Block Diagonal} \left(h_1(\theta), h_2(\theta) \right),$$

where

$$h_1(\theta) = \frac{\theta_2^2 (c_{11}c_{22} - c_{12}^2)}{\theta_{r+1}^2 (c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22})^2}, \quad \text{and}$$

$$h_2(\theta) = \theta_{r+1}^{-2} \text{Block Diagonal} \left(1, ((s_{ij})), 2n \right), \quad i, j = 3, \dots, r.$$

Hence, from Theorem 1 of Datta and M.Ghosh (1995), the two-group reference prior is given by

$$\pi^{2R}(\boldsymbol{\theta}) \propto \frac{1}{\theta_{r+1}^r} \left(c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22} \right)^{-1}. \quad (4.17)$$

Next we consider the three-group reference prior. The parameter space is partitioned into $\{\theta_1\}, \{\theta_2\}, \{\theta_3, \dots, \theta_{r+1}\}$, θ_1 being still the parameter of interest. Let $\boldsymbol{\theta}_{(1)} = \theta_1$, $\boldsymbol{\theta}_{(2)} = (\theta_2)$ and $\boldsymbol{\theta}_{(3)} = (\theta_3, \dots, \theta_{r+1})$. Then the Fisher information matrix can be rewritten as

$$\mathbf{I}(\boldsymbol{\theta}) = \text{Block Diagonal} \left(h_1(\boldsymbol{\theta}), h_2(\boldsymbol{\theta}), h_3(\boldsymbol{\theta}) \right),$$

where

$$h_1(\boldsymbol{\theta}) = \frac{\theta_2^2(c_{11}c_{22} - c_{12}^2)}{\theta_{r+1}^2(c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22})^2},$$

$$h_2(\boldsymbol{\theta}) = \theta_{r+1}^{-2}, \text{ and}$$

$$h_3(\boldsymbol{\theta}) = \theta_{r+1}^{-2} \text{Block Diagonal} \left((s_{jl}), 2n \right), \quad j, l = 3, \dots, r.$$

From Theorem 1 of Datta and M.Ghosh (1995) and following similar steps to the previous case, the three-group reference prior is given by

$$\pi^{3R}(\boldsymbol{\theta}) \propto \frac{1}{\theta_{r+1}^{r-1}} \left(c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22} \right)^{-1}. \quad (4.18)$$

A second three group reference prior with the partition $\{\theta_1\}, \{\theta_2, \dots, \theta_r\}, \{\theta_{r+1}\}$ leads to the three-group reference prior $\theta_{r+1}^{-1}(c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22})^{-1}$.

Finally, the one-at-a-time reference prior with the partition $\{\theta_1\}, \{\theta_2\}, \dots, \{\theta_r\}, \{\theta_{r+1}\}$ where θ_1 is the parameter of interest, while the remaining parameters are considered in any arbitrary order, is given by the same reference prior $\theta_{r+1}^{-1}(c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22})^{-1}$.

4.4.3 Probability Matching Priors

Due to the orthogonality of θ_1 with $(\theta_2, \dots, \theta_{r+1})$ (cf. Cox and Reid, 1987), from Tibshirani (1989), the class of first order probability matching priors is characterized by

$$\pi^F(\theta) \propto \frac{|\theta_2|}{\theta_{r+1}} \left(c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22} \right)^{-1} q(\theta_2, \dots, \theta_{r+1}) , \quad (4.19)$$

where q is any arbitrary positive-valued function differentiable in its arguments. Jeffreys' prior is a first order probability matching prior and the other reference priors are all first order probability matching priors. Indeed, there are infinitely many first order probability matching prior. To narrow down the selection of priors within this class, we now consider second order probability matching priors.

From Mukerjee and Ghosh (1997), due to the orthogonality of θ_1 with $(\theta_2, \dots, \theta_{r+1})$, this class of priors is characterized as solutions of the second partial differential equation

$$\frac{1}{6}q(\theta_2, \dots, \theta_{r+1}) \frac{\partial}{\partial \theta_1} (I_{11}^{-3/2} L_{1,1,1}) + \sum_{v=2}^{r+1} \sum_{s=2}^{r+1} \frac{\partial}{\partial \theta_v} \{ I_{11}^{-1/2} L_{11s} I^{sv} q(\theta_2, \dots, \theta_{r+1}) \} = 0 , \quad (4.20)$$

where $L_{1,1,1} = E \left[\frac{\partial \log L}{\partial \theta_1} \right]^3 = 0$. Also,

$$\begin{aligned} \frac{\partial \log L}{\partial \theta_1} &= \theta_{r+1}^{-2} \sum_{i=1}^n \left\{ [y_i - \theta_2 h(\theta_1)] \{ \theta_1 x_{i1} + x_{i2} - \sum_{j=3}^r x_{ij} (a_{j1} \theta_1 + a_{j2}) \} - \sum_{j=3}^r x_{ij} \theta_j \right\} \\ &\quad \times \left\{ \theta_2 [h'(\theta_1)] \{ \theta_1 x_{i1} + x_{i2} - \sum_{j=3}^r x_{ij} (a_{j1} \theta_1 + a_{j2}) \} + h(\theta_1) (x_{i1} - \sum_{j=3}^r x_{ij} a_{j1}) \right\} . \end{aligned}$$

For $k = 3, \dots, r$,

$$\frac{\partial^2 \log L}{\partial \theta_1 \partial \theta_k} = \frac{\theta_2}{\theta_{r+1}^2} \sum_{i=1}^n x_{ik} \left[h'(\theta_1) \{ \theta_1 x_{i1} + x_{i2} - \sum_{j=3}^r x_{ij} (a_{j1} \theta_1 + a_{j2}) \} + h(\theta_1) (x_{i1} - \sum_{j=3}^r x_{ij} a_{j1}) \right]$$

$$\begin{aligned}
&= \frac{\theta_2}{\theta_{r+1}^2} \left[h'(\theta_1) \{ \theta_1 (s_{k1} - \sum_{j=3}^r s_{kj} a_{j1}) + s_{k2} - \sum_{j=3}^r s_{kj} a_{j2} \} + h(\theta_1) \{ s_{k1} - \sum_{j=3}^r s_{kj} a_{j1} \} \right] \\
&= \frac{\theta_2}{\theta_{r+1}^2} \left[h'(\theta_1) \{ \theta_1 \cdot 0 + 0 \} + h(\theta_1) \cdot 0 \right] \\
&= 0
\end{aligned}$$

by the definition of s_{k1} and s_{k2} for $k = 3, \dots, r$. Hence

$$\frac{\partial^3 \log L}{\partial \theta_1^2 \partial \theta_k} = 0, \quad k = 3, \dots, r.$$

Thus,

$$L_{11k} = E \left[\frac{\partial^3 \log L}{\partial \theta_1^2 \partial \theta_k} \right] = 0, \quad k = 3, \dots, r.$$

Also, since

$$\frac{\partial^2 \log L}{\partial \theta_1 \partial \theta_{r+1}} = -\frac{2}{\theta_{r+1}} \frac{\partial \log L}{\partial \theta_1},$$

and

$$\frac{\partial^3 \log L}{\partial \theta_1^2 \partial \theta_{r+1}} = -\frac{2}{\theta_{r+1}} \frac{\partial^2 \log L}{\partial \theta_1^2},$$

one has,

$$L_{11\overline{r+1}} = E \left[\frac{\partial^3 \log L}{\partial \theta_1^2 \partial \theta_{r+1}} \right] = \frac{2}{\theta_{r+1}} I_{11}.$$

Next

$$\begin{aligned}
\frac{\partial^2 \log L}{\partial \theta_1 \partial \theta_2} &= -\frac{1}{\theta_{r+1}^2} \sum_{i=1}^n \left\{ [y_i - \theta_2 h(\theta_1)] \{ \theta_1 x_{i1} + x_{i2} - \sum_{j=3}^r x_{ij} (a_{j1} \theta_1 + a_{j2}) \} - \sum_{j=3}^r x_{ij} \theta_j \right. \\
&\quad \times [h'(\theta_1) \{ \theta_1 x_{i1} + x_{i2} - \sum_{j=3}^r x_{ij} (a_{j1} \theta_1 + a_{j2}) \} + h(\theta_1) (x_{i1} - \sum_{j=3}^r x_{ij} a_{j1})] \Big\} \\
&\quad + \frac{\theta_2}{\theta_{r+1}^2} h(\theta_1) \sum_{i=1}^n \left\{ [\theta_1 x_{i1} + x_{i2} - \sum_{j=3}^r x_{ij} (a_{j1} \theta_1 + a_{j2})] \right.
\end{aligned}$$

$$\times [h'(\theta_1)\{\theta_1 x_{i1} + x_{i2} - \sum_{j=3}^r x_{ij}(a_{j1}\theta_1 + a_{j2})\} + h(\theta_1)(x_{i1} - \sum_{j=3}^r x_{ij}a_{j1})]\Big\}$$

After some algebra, we get

$$L_{112} = \frac{I_{11}}{\theta_2} \quad .$$

Hence, (4.20) simplifies to

$$\sum_{v=2}^{r+1} \frac{\partial}{\partial \theta_v} \left[\theta_2^{-1} (I_{11}^{1/2} I^{2v} + \frac{2}{\theta_{r+1}} I_{11}^{1/2} I^{r+1,v}) q(\theta_2, \dots, \theta_{r+1}) \right] = 0 \quad (4.21)$$

This simplifies to

$$\frac{\partial}{\partial \theta_2} \left[\theta_{r+1} \frac{|\theta_2|}{\theta_2} q(\theta_2, \dots, \theta_{r+1}) \right] + \frac{\partial}{\partial \theta_{r+1}} \left[\frac{|\theta_2|}{2n} q(\theta_2, \dots, \theta_{r+1}) \right] = 0 \quad (4.22)$$

or equivalently

$$\theta_{r+1} (\text{sgn}[\theta_2]) \frac{\partial q}{\partial \theta_2} + \frac{|\theta_2|}{2n} \frac{\partial q}{\partial \theta_{r+1}} = 0 \quad .$$

Clearly $q(\theta_2, \dots, \theta_{r+1}) \propto 1$ provides a solution to (4.22). Thus, a second order probability matching prior is given by

$$\pi^s(\theta) \propto \frac{|\theta_2|}{\theta_{r+1}} \left(c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22} \right)^{-1} \quad (4.23)$$

4.5 Propriety of Posteriors

In this section, we prove the propriety of posteriors under the priors developed in the previous section.

Theorem 4.2 Under the class of priors $\pi(\theta) \propto \theta_{r+1}^{-a} Q^{-1}(\theta_1)$, the marginal posterior distribution of θ_1 is proper when $n + a > r + 1$, and is given by

$$\pi(\theta_1|\mathbf{y}) \propto Q^{-1}(\theta_1) \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} \right\}^{-\frac{n+a-r+1}{2}},$$

where

$$Q(\theta_1) = (c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22}),$$

$$SSE = \sum_{i=1}^n (y_i - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_r x_{ir})^2,$$

and

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{bmatrix}.$$

The proof of this theorem is deferred to the appendix.

Remark 4.1 This class of priors includes the two-group, three-group and one-at-a-time reference priors with $a = r$, $r - 1$ and 1 , respectively.

Theorem 4.3 Under the class of priors $\pi(\theta) \propto \frac{|\theta_2|}{\theta_1^a} Q^{-1}(\theta_1)$, the marginal posterior distribution of θ_1 is proper when $n + a > r + 3$, and is given by

$$\begin{aligned} & \frac{1}{Q^{\frac{1}{2}}(\theta_1)} \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} + Q(\theta_1)\omega^2(\theta_1) \right\}^{-\frac{n+a-r+3}{2}} \\ & + 2\omega(\theta_1) \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} \right\}^{-\frac{n+a-r}{2}} \int_0^A \frac{dz}{(1+z^2)^{\frac{n+a-r+1}{2}}}, \end{aligned}$$

where

$$\omega(\theta_1) = \frac{\theta_1(c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2) + c_{12}\hat{\beta}_1 + c_{22}\hat{\beta}_2}{Q(\theta_1)}$$

$$A = Q^{\frac{1}{2}}(\theta_1)\omega(\theta_1)\sqrt{SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)}}.$$

The proof of this theorem is also deferred to the appendix.

Remark 4.2 In Theorem 4.3 , $a = r + 1$ and 1 correspond to the Jeffreys' prior and second order probability matching prior, respectively.

Remark 4.3 The prior used by Hunter and Lamboy (1981) for linear calibration problem, $\pi_{HL} \propto |\beta|\sigma^{-2}$, is a second order probability matching prior. The general class of first order probability matching priors for linear calibration problem was also derived in Ghosh, Carlin and Srivastava (1995).

Remark 4.4 Ghosh and Yin (2000) derived probability matching priors and reference priors for the generalized Fieller-Creasy problem which involved inference about the ratio of two location parameters for two independent symmetric location-scale distribution.

Remark 4.5 The two-group reference prior was derived by Mendoza (1990) for slope-ratio bioassay problem in the normal case.

Remark 4.6 Kim, Carter and Hubert (1991) derived a closed-form point and interval estimator in a symmetric parallel-line bioassay.

Remark 4.7 Buonaccorsi and Gatsonis (1988) provided a Bayesian solution to the problem of inference about the ratio of two coefficients in a linear model. But, they did not provide a way of choosing the function $h(\cdot)$.

Now, the resulting posteriors for these special cases are characterized by finding the elements c_{11} , c_{12} and c_{22} of C using Theorem 4.2 and 4.3. Table 4.1 gives the all elements of C obtained from the design matrix X for the special cases.

Table 4.1: Catalog of the elements of C

Cases	Model	c_{11} , c_{12} and c_{22}
Linear Calibration Problem	(4.3)	$c_{11} = \frac{mk}{m+k}$ $c_{12} = -\frac{k}{m+k} \sum_{i=1}^m x_i$ $c_{22} = \sum_{i=1}^m x_i^2 - \frac{1}{m+k} \left[\sum_{i=1}^m x_i \right]^2$
Fieller Creasy Problem	(4.5)	$c_{11} = k$ $c_{12} = 0$ $c_{22} = m$
Slope-Ratio Bioassay Problem	(4.7)	$c_{11} = u \sum_{j=1}^q x_{2j}^2 - \frac{u^2}{pm+qu} \left[\sum_{j=1}^q x_{2j} \right]^2$ $c_{12} = -\frac{mu}{pm+qu} \left[\sum_{i=1}^p x_{1i} \right] \left[\sum_{j=1}^q x_{2j} \right]$ $c_{22} = m \sum_{i=1}^p x_{1i}^2 - \frac{m^2}{pm+qu} \left[\sum_{i=1}^p x_{1i} \right]^2$
Parallel-Line Bioassay Problem	(4.11)	$c_{11} = \frac{pmqu}{pm+qu}$ $c_{12} = u \sum_{j=1}^q x_{2j} - \frac{qu}{pm+qu} \left[\sum_{i=1}^p x_{1i} + \sum_{j=1}^q x_{2j} \right]$ $c_{22} = m \sum_{i=1}^p x_{1i}^2 + u \sum_{j=1}^q x_{2j}^2 - \frac{1}{pm+qu} \left[m \sum_{i=1}^p x_{1i} + u \sum_{j=1}^q x_{2j} \right]^2$

4.6 Numerical Examples

In order to see the performance of the noninformative priors derived in Section 4.4, we analyze two sets of data using Jeffreys' prior, the reference priors and second order probability matching prior.

Example 1 Parallel-Line Bioassay Problem

Finney (1978, p105) analyzed the data in Table 4.2 from turbidimetric measurements on the growth response of *Lactobacillus leichmannii* to vitamin B_{12} which is reported by Emery et al.(1951).

Table 4.2: Responses in an assay of vitamin B_{12}

Stimulus	Standard			Test		
Dose	-1.0	0.0	1.0	-1.0	0.0	1.0
	0.96	1.06	1.17	0.91	1.09	1.15
	0.91	1.07	1.14	0.93	1.04	1.15
	0.92	0.99	1.14	0.98	0.97	1.14
	0.76	0.86	1.13	0.96	1.06	1.16
	1.03	1.06	1.13	0.89	1.04	1.10
	0.93	1.02	1.15	1.01	1.02	1.15

In this case, $n = 36, r = 3, p = 3, q = 3, SSE = 0.0961$ and

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \hat{\beta}\rho \\ \hat{\beta} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} 0.0178 \\ 0.105 \\ 1.0238 \end{bmatrix}, \quad C = \begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{bmatrix} = \begin{bmatrix} 9 & 0 \\ 0 & 24 \end{bmatrix}.$$

Finney obtained Fieller's 95% confidence interval for this data as $(-0.193, 0.551)$ and the length of interval is 0.744. Table 4.3 shows that the noninformatives priors performed well with shorter intervals than Finney's.

Table 4.3: Posterior quantiles

	$P_{0.025}$	$P_{0.975}$	Length
π^J	-0.165	0.523	0.688
π^{1R}	-0.187	0.536	0.723
π^{2R}	-0.176	0.523	0.699
π^{3R}	-0.181	0.529	0.710
π^S	-0.182	0.541	0.723

Example 2. Slope-Ratio Bioassay Problem

The data in Table 4.4 were reported by Wood (1946), who used 20 tubes, 4 each for blanks, 0.1, 0.2 μg of standard riboflavin and 0.025, 0.5 g malt per tube. Responses were measured to the nearest 0.05 ml : for arithmetic convenience, this may be taken as the unit of responses, so that all values of y are integers. It was analyzed by Finney (1978, p161) using the classical approach.

Table 4.4: Responses in an assay of riboflavin in malt

Stimulus Dose	Standard			Test	
	0.0	0.5	1.0	0.5	1.0
	38	97	167	80	121
	45	100	164	88	124
	40	105	159	90	122
	44	98	156	82	122

In this case, $n = 20$, $r = 3$, $p = 3$, $q = 2$, $SSE = 252.8857$ and

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} \hat{\beta}\rho \\ \hat{\beta} \\ \hat{\alpha} \end{bmatrix} = \begin{bmatrix} 81.2288 \\ 118.6285 \\ 42.1429 \end{bmatrix}, \quad C = \begin{bmatrix} c_{11} & c_{12} \\ c_{12} & c_{22} \end{bmatrix} = \begin{bmatrix} 3.2 & -1.8 \\ -1.8 & 3.2 \end{bmatrix}.$$

For this dataset, the 95% confidence interval obtained by Finney was (0.6464, 0.7235) and the length of interval is 0.0771. On the other hand, from Table 4.5 we can see that the 0.025th and 0.975th percent quantile points corresponding to Jeffreys prior (π^J), one-at-a-time reference prior (π^{1R}), two-group reference prior (π^{2R}), three-group reference prior (π^{3R}), and second order probability matching prior (π^S) are similar or shorter than Finney's result.

Table 4.5: Posterior quantiles

	$P_{0.025}$	$P_{0.975}$	Length
π^J	0.6504	0.7195	0.0691
π^{1R}	0.6460	0.7241	0.0781
π^{2R}	0.6486	0.7218	0.0732
π^{3R}	0.6473	0.7227	0.0754
π^S	0.6469	0.7227	0.0758

4.7 Appendix

Proof of Theorem 4.1 Let $\mathbf{I}(\boldsymbol{\theta}) = \mathbf{J}\mathbf{I}(\boldsymbol{\beta}, \sigma)\mathbf{J}^T$. Then

$$(\mathbf{J}\mathbf{I})_{11} = \frac{\theta_2}{\theta_{r+1}^2} \left[s_{11} \{ \theta_1 h'(\theta_1) + h(\theta_1) \} + s_{12} h'(\theta_1) - \sum_{j=3}^r s_{1j} g'_j(\theta_1) \right]$$

$$(\mathbf{J}\mathbf{I})_{12} = \frac{\theta_2}{\theta_{r+1}^2} \left[s_{12} \{ \theta_1 h'(\theta_1) + h(\theta_1) \} + s_{22} h'(\theta_1) - \sum_{j=3}^r s_{2j} g'_j(\theta_1) \right]$$

$$(\mathbf{J}\mathbf{I})_{1i} = \frac{\theta_2}{\theta_{r+1}^2} \left[s_{1i} \{ \theta_1 h'(\theta_1) + h(\theta_1) \} + s_{2i} h'(\theta_1) - \sum_{j=3}^r s_{ij} g'_j(\theta_1) \right], \quad i = 3, \dots, r$$

$$(\mathbf{J}\mathbf{I})_{1,r+1} = 0$$

Hence, if

$$\begin{bmatrix} g'_3(\theta_1) \\ \vdots \\ g'_r(\theta_1) \end{bmatrix} = \begin{bmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{3r} & \cdots & s_{rr} \end{bmatrix}^{-1} \begin{bmatrix} s_{13} & s_{23} \\ \vdots & \vdots \\ s_{1r} & s_{2r} \end{bmatrix} \begin{bmatrix} \theta_1 h'(\theta_1) + h(\theta_1) \\ h'(\theta_1) \end{bmatrix},$$

then, one has $(\mathbf{J}\mathbf{I})_{1i} = 0$, $i = 3, \dots, r$. So pick

$$\begin{aligned} \begin{bmatrix} g_3(\theta_1) \\ \vdots \\ g_r(\theta_1) \end{bmatrix} &= \begin{bmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{3r} & \cdots & s_{rr} \end{bmatrix}^{-1} \begin{bmatrix} s_{13} & s_{23} \\ \vdots & \vdots \\ s_{1r} & s_{2r} \end{bmatrix} \begin{bmatrix} \theta_1 h(\theta_1) \\ h(\theta_1) \end{bmatrix} \\ &= \begin{bmatrix} a_{31} & a_{32} \\ \vdots & \vdots \\ a_{r1} & a_{r2} \end{bmatrix} \begin{bmatrix} \theta_1 h(\theta_1) \\ h(\theta_1) \end{bmatrix} \\ &= h(\theta_1) \begin{bmatrix} a_{31}\theta_1 + a_{32} \\ \vdots \\ a_{r1}\theta_1 + a_{r2} \end{bmatrix} \end{aligned}$$

Now, since $\mathbf{I}(\theta) = \mathbf{J}\mathbf{I}(\beta, \sigma)\mathbf{J}^T$, one has

$$(\mathbf{I}(\theta))_{1i} = (\mathbf{J}\mathbf{I})_{1i} = 0, \quad i = 3, \dots, r, r+1.$$

Similarly,

$$(\mathbf{I}(\theta))_{2i} = (\mathbf{J}\mathbf{I})_{2i} = 0, \quad i = 3, \dots, r, r+1.$$

Note that

$$\begin{aligned} (\mathbf{I}(\theta))_{12} &= (\mathbf{J}\mathbf{I}\mathbf{J}^T)_{12} \\ &= \frac{\theta_2 h(\theta_1)}{\theta_{r+1}^2} \left[s_{11}\theta_1 \{ \theta_1 h'(\theta_1) + h(\theta_1) \} + 2s_{12}\theta_1 h'(\theta_1) - \theta_1 \sum_{j=3}^r s_{1j} g'_j(\theta_1) \right. \\ &\quad \left. + s_{12}h(\theta_1) + s_{22}h'(\theta_1) - \sum_{j=3}^r s_{2j} g'_j(\theta_1) \right] \\ &= \frac{\theta_2 h(\theta_1)}{\theta_{r+1}^2} \left[s_{11}\theta_1 \{ \theta_1 h'(\theta_1) + h(\theta_1) \} + 2s_{12}\theta_1 h'(\theta_1) \right. \end{aligned}$$

$$\begin{aligned}
& -\theta_1 \sum_{j=3}^r s_{1j} \{a_{j1} [\theta_1 h'(\theta_1) + h(\theta_1)] + a_{j2} h'(\theta_1)\} \\
& + s_{12} h(\theta_1) + s_{22} h'(\theta_1) - \sum_{j=3}^r s_{2j} \{a_{j1} [\theta_1 h'(\theta_1) + h(\theta_1)] + a_{j2} h'(\theta_1)\} \Big]
\end{aligned}$$

We need $I_{12} = 0$ to satisfy the condition of orthogonality. Thus, we require

$$\begin{aligned}
0 = & h'(\theta_1) \left[\theta_1^2 (s_{11} - \sum_{j=3}^r s_{1j} a_{j1}) + \theta_1 (2s_{12} - \sum_{j=3}^r s_{1j} a_{j2} - \sum_{j=3}^r s_{2j} a_{j1}) \right. \\
& \left. + s_{22} - \sum_{j=3}^r s_{2j} a_{j2} \right] + h(\theta_1) \left[\theta_1 (s_{11} - \sum_{j=3}^r s_{1j} a_{j1}) + s_{12} - \sum_{j=3}^r s_{2j} a_{j1} \right] .
\end{aligned}$$

Also $\sum_{j=3}^r s_{1j} a_{j2} = \sum_{j=3}^r s_{2j} a_{j1}$, since

$$\begin{aligned}
\sum_{j=3}^r s_{1j} a_{j2} &= (s_{13} \ \cdots \ s_{1r}) \begin{bmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{3r} & \cdots & s_{rr} \end{bmatrix}^{-1} \begin{bmatrix} s_{23} \\ \vdots \\ s_{2r} \end{bmatrix} \\
\sum_{j=3}^r s_{2j} a_{j1} &= (s_{23} \ \cdots \ s_{2r}) \begin{bmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{3r} & \cdots & s_{rr} \end{bmatrix}^{-1} \begin{bmatrix} s_{13} \\ \vdots \\ s_{1r} \end{bmatrix} .
\end{aligned}$$

Hence, one needs to find $h(\theta_1)$ from

$$\begin{aligned}
0 = & h'(\theta_1) \left[\theta_1^2 (s_{11} - \sum_{j=3}^r s_{1j} a_{j1}) + 2\theta_1 (s_{12} - \sum_{j=3}^r s_{1j} a_{j2}) + s_{22} - \sum_{j=3}^r s_{2j} a_{j2} \right] \\
& + h(\theta_1) \left[\theta_1 (s_{11} - \sum_{j=3}^r s_{1j} a_{j1}) + s_{12} - \sum_{j=3}^r s_{1j} a_{j2} \right] .
\end{aligned}$$

Thus,

$$\frac{h'(\theta_1)}{h(\theta_1)} = - \frac{\theta_1 (s_{11} - \sum_{j=3}^r s_{1j} a_{j1}) + s_{12} - \sum_{j=3}^r s_{1j} a_{j2}}{\theta_1^2 (s_{11} - \sum_{j=3}^r s_{1j} a_{j1}) + 2\theta_1 (s_{12} - \sum_{j=3}^r s_{1j} a_{j2}) + s_{22} - \sum_{j=3}^r s_{2j} a_{j2}}$$

or equivalently

$$\log h(\theta_1) = -\frac{1}{2} \left[\theta_1^2 (s_{11} - \sum_{j=3}^r s_{1j} a_{j1}) + 2\theta_1 (s_{12} - \sum_{j=3}^r s_{1j} a_{j2}) + s_{22} - \sum_{j=3}^r s_{2j} a_{j2} \right] .$$

This leads to

$$\begin{aligned} h(\theta_1) &= \left[\theta_1^2 (s_{11} - \sum_{j=3}^r s_{1j} a_{j1}) + 2\theta_1 (s_{12} - \sum_{j=3}^r s_{1j} a_{j2}) + s_{22} - \sum_{j=3}^r s_{2j} a_{j2} \right]^{-\frac{1}{2}} \\ &= \left[c_{11}\theta_1^2 + 2c_{12}\theta_1 s_{12} + c_{22} \right]^{-\frac{1}{2}} \end{aligned}$$

Proof of Theorem 4.2 Recall that

$$\begin{aligned} \beta_2 &= \theta_2 h(\theta_1) \quad , \\ \beta_j &= \theta_j - \theta_2 h(\theta_1) (a_{j1}\theta_1 + a_{j2}) \\ &= \theta_j - \beta_2 (a_{j1}\theta_1 + a_{j2}) \quad , \quad j = 3, \dots, r \\ \sigma &= \theta_{r+1} \end{aligned}$$

Then,

$$\begin{aligned} \theta_1 &= \theta_1 \quad , \\ \theta_2 &= \frac{\beta_2}{h(\theta_1)} \quad , \\ \theta_j &= \beta_j + \beta_2 (a_{j1}\theta_1 + a_{j2}) \quad , \quad j = 3, \dots, r \\ \theta_{r+1} &= \sigma \quad . \end{aligned}$$

Hence, the Jacobian of transformation is $h^{-1}(\theta_1) = (c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22})^{\frac{1}{2}}$. Now,

$$\pi(\theta_1, \beta_2, \dots, \beta_r, \sigma) \propto \sigma^{-a} (c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22})^{-\frac{1}{2}} \quad .$$

Thus the likelihood function is

$$L(\theta_1, \beta_2, \dots, \beta_r, \sigma) \propto \sigma^{-n} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \left[y_i - \beta_2 \theta_1 x_{i1} - \beta_2 x_{i2} - \sum_{j=3}^r \beta_j x_{ij} \right]^2 \right]$$

$$= \sigma^{-n} \exp \left[-\frac{1}{2\sigma^2} \left\{ SSE + \beta_*^T \mathbf{X}^T \mathbf{X} \beta_* \right\} \right],$$

where $\beta_*^T = (\beta_2\theta_1 - \hat{\beta}_1, \beta_2 - \hat{\beta}_2, \dots, \beta_r - \hat{\beta}_r)$. Hence the joint posterior of $\theta_1, \beta_2, \dots, \beta_r$ and σ is given by

$$\pi(\theta_1, \beta_2, \dots, \beta_r, \sigma | \mathbf{y}) \propto \sigma^{-(n+a)} \exp \left[-\frac{1}{2\sigma^2} \left\{ SSE + \beta_*^T \mathbf{X}^T \mathbf{X} \beta_* \right\} \right] (c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22})^{-\frac{1}{2}}.$$

Write

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} s_{11} & s_{12} & s_{13} & \cdots & s_{1r} \\ s_{21} & s_{22} & s_{23} & \cdots & s_{2r} \\ s_{31} & s_{32} & s_{33} & \cdots & s_{3r} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{r1} & s_{r2} & s_{r3} & \cdots & s_{rr} \end{bmatrix}$$

Since

$$\begin{aligned} \mathbf{C} &= \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \\ &= \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} - \begin{bmatrix} s_{13} & \cdots & s_{1r} \\ s_{23} & \cdots & s_{2r} \end{bmatrix} \begin{bmatrix} s_{33} & \cdots & s_{3r} \\ \vdots & \ddots & \vdots \\ s_{r3} & \cdots & s_{rr} \end{bmatrix}^{-1} \begin{bmatrix} s_{13} & s_{23} \\ \vdots & \vdots \\ s_{1r} & s_{2r} \end{bmatrix} \end{aligned}$$

integrating with respect to $\beta_3, \dots, \beta_r$, and using the properties of multivariate normal distribution, one gets

$$\begin{aligned} \pi(\theta_1, \beta_2, \sigma | \mathbf{y}) &\propto \sigma^{-(n+a)+r-2} Q^{-\frac{1}{2}}(\theta_1) \\ &\times \exp \left[-\frac{1}{2\sigma^2} \left\{ SSE + c_{11}(\beta_2\theta_1 - \hat{\beta}_1)^2 + 2c_{12}(\beta_2\theta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + c_{22}(\beta_2 - \hat{\beta}_2)^2 \right\} \right] \end{aligned}$$

Next write

$$\begin{aligned}
& c_{11}(\beta_2\theta_1 - \hat{\beta}_1)^2 + 2c_{12}(\beta_2\theta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + c_{22}(\beta_2 - \hat{\beta}_2)^2 \\
&= \beta_2^2 Q(\theta_1) - 2\beta_2[\theta_1(c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2) + (c_{12}\hat{\beta}_1 + c_{22}\hat{\beta}_2)] + c_{11}\hat{\beta}_1^2 + 2c_{12}\hat{\beta}_1\hat{\beta}_2 + c_{22}\hat{\beta}_2^2 \\
&= Q(\theta_1) \left\{ \beta_2 - \frac{\theta_1(c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2) + (c_{12}\hat{\beta}_1 + c_{22}\hat{\beta}_2)}{Q(\theta_1)} \right\}^2 \\
&\quad + c_{11}\hat{\beta}_1^2 + 2c_{12}\hat{\beta}_1\hat{\beta}_2 + c_{22}\hat{\beta}_2^2 - \frac{\left\{ \theta_1(c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2) + (c_{12}\hat{\beta}_1 + c_{22}\hat{\beta}_2) \right\}^2}{Q(\theta_1)}
\end{aligned}$$

But on simplification,

$$\begin{aligned}
& c_{11}\hat{\beta}_1^2 + 2c_{12}\hat{\beta}_1\hat{\beta}_2 + c_{22}\hat{\beta}_2^2 - \frac{\left\{ \theta_1(c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2) + (c_{12}\hat{\beta}_1 + c_{22}\hat{\beta}_2) \right\}^2}{Q(\theta_1)} \\
&= Q^{-1}(\theta_1) \left\{ (c_{11}\hat{\beta}_1^2 + 2c_{12}\hat{\beta}_1\hat{\beta}_2 + c_{22}\hat{\beta}_2^2)(c_{11}\theta_1^2 + 2c_{12}\theta_1 + c_{22}) \right. \\
&\quad \left. - \left[\theta_1(c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2) + c_{12}\hat{\beta}_1 + c_{22}\hat{\beta}_2 \right]^2 \right\} \\
&= Q^{-1}(\theta_1) |C| (\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2
\end{aligned}$$

Hence, integrating with respect to β_2 ,

$$\pi(\theta_1, \sigma | \mathbf{y}) \propto \sigma^{-(n+a)+r-1} Q(\theta_1) \exp \left[-\frac{1}{2\sigma^2} \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} \right\} \right].$$

Finally, integrating with respect to σ ,

$$\pi(\theta_1 | \mathbf{y}) \propto Q^{-1}(\theta_1) \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} \right\}^{-\frac{n+a-r+1}{2}},$$

Now observe that since $n + a > r$,

$$\pi(\theta_1|\mathbf{y}) \leq KQ^{-1}(\theta_1) \quad ,$$

where $K(> 0)$ is a generic constant depending on $\mathbf{y}$, n , r , and a but not on θ_1 .

Since,

$$Q^{-1}(\theta_1) = \left\{ c_{11} \left[\theta_1 + \frac{c_{12}}{c_{11}} \right]^2 + c_{22} - \frac{c_{12}^2}{c_{11}} \right\}^{-1}$$

and $\mathbf{C}$ is positive definite so that

$$c_{22} - \frac{c_{12}^2}{c_{11}} > 0, \quad Q^{-1}(\theta_1) \leq c_{11}^{-1} \left[\theta_1 + \frac{c_{12}}{c_{11}} \right]^{-2} \quad .$$

Hence, from the property of the Cauchy distribution, the finiteness of $\int_{-\infty}^{\infty} \pi(\theta_1|\mathbf{y}) d\theta_1$ follows.

Proof of Theorem 4.3 Since $\theta_2 = \frac{\beta_2}{h(\theta_1)}$, proceeding as in the previous case, integrating with respect to $\beta_3, \dots, \beta_r$ one gets

$$\begin{aligned} \pi(\theta_1, \beta_2, \sigma|\mathbf{y}) &\propto \sigma^{-(n+a)+r-2} |\beta_2| Q^{-\frac{1}{2}}(\theta_1) \\ &\times \exp \left[-\frac{1}{2\sigma^2} \left\{ SSE + c_{11}(\beta_2\theta_1 - \hat{\beta}_1)^2 + 2c_{12}(\beta_2\theta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + c_{22}(\beta_2 - \hat{\beta}_2)^2 \right\} \right] \end{aligned}$$

First, integrating with respect to σ

$$\begin{aligned} \pi(\theta_1, \beta_2|\mathbf{y}) &\propto |\beta_2| Q^{-\frac{1}{2}}(\theta_1) \\ &\times \left\{ SSE + c_{11}(\beta_2\theta_1 - \hat{\beta}_1)^2 + 2c_{12}(\beta_2\theta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + c_{22}(\beta_2 - \hat{\beta}_2)^2 \right\}^{-\frac{n+a-r+1}{2}} \end{aligned}$$

As before, we simplify

$$\begin{aligned} &c_{11}(\beta_2\theta_1 - \hat{\beta}_1)^2 + 2c_{12}(\beta_2\theta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + c_{22}(\beta_2 - \hat{\beta}_2)^2 \\ &= Q(\theta_1) \left\{ \beta_2 - \frac{\theta_1(c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2) + (c_{12}\hat{\beta}_1 + c_{22}\hat{\beta}_2)}{Q(\theta_1)} \right\}^2 + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} \end{aligned}$$

Now, find

$$\begin{aligned}
& \int_{-\infty}^{\infty} |\beta_2| \left\{ SSE + Q(\theta_1) [\beta_2 - \omega(\theta_1)]^2 + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} \right\}^{-\frac{n+a-r+1}{2}} d\beta_2 \\
& \quad , \text{ where } \omega(\theta_1) = \frac{\theta_1(c_{11}\hat{\beta}_1 + c_{12}\hat{\beta}_2) + c_{12}\hat{\beta}_1 + c_{22}\hat{\beta}_2}{Q(\theta_1)} \\
& = \int_0^{\infty} \beta_2 \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} + Q(\theta_1) [\beta_2 - \omega(\theta_1)]^2 \right\}^{-\frac{n+a-r+1}{2}} d\beta_2 \\
& - \int_{-\infty}^0 \beta_2 \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} + Q(\theta_1) [\beta_2 - \omega(\theta_1)]^2 \right\}^{-\frac{n+a-r+1}{2}} d\beta_2
\end{aligned}$$

Let

$$z = Q^{\frac{1}{2}}(\theta_1) [\beta_2 - \omega(\theta_1)] .$$

Then

$$\begin{aligned}
\text{1st Term} &= \frac{1}{Q^{\frac{1}{2}}(\theta_1)} \int_{-Q^{\frac{1}{2}}(\theta_1)\omega(\theta_1)}^{\infty} \frac{z}{\left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} + z^2 \right\}^{\frac{n+a-r+1}{2}}} dz \\
&+ \omega(\theta_1) \int_{-Q^{\frac{1}{2}}(\theta_1)\omega(\theta_1)}^{\infty} \frac{dz}{\left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} + z^2 \right\}^{\frac{n+a-r+1}{2}}} \\
&= \frac{1}{Q^{\frac{1}{2}}(\theta_1)} \frac{1}{2} \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} + Q(\theta_1)\omega^2(\theta_1) \right\}^{-\frac{n+a-r+3}{2}} \\
&+ \omega(\theta_1) \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} \right\}^{-\frac{n+a-r}{2}} \int_{-A}^{\infty} \frac{dz}{(1+z^2)^{\frac{n+a-r+1}{2}}} ,
\end{aligned}$$

where

$$A = Q^{\frac{1}{2}}(\theta_1)\omega(\theta_1) \sqrt{SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)}} .$$

Similarly,

$$\begin{aligned} \text{2nd Term} &= \frac{1}{Q^{\frac{1}{2}}(\theta_1)} \frac{1}{2} \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} + Q(\theta_1)\omega^2(\theta_1) \right\}^{-\frac{n+a-r+3}{2}} \\ &\quad - \omega(\theta_1) \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} \right\}^{-\frac{n+a-r}{2}} \int_A \frac{dz}{(1+z^2)^{\frac{n+a-r+1}{2}}} . \end{aligned}$$

Thus, the marginal posterior pdf of θ_1 is given by

$$\begin{aligned} &\frac{1}{Q^{\frac{1}{2}}(\theta_1)} \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} + Q(\theta_1)\omega^2(\theta_1) \right\}^{-\frac{n+a-r+3}{2}} \\ &+ 2\omega(\theta_1) \left\{ SSE + \frac{|C|(\hat{\beta}_2\theta_1 - \hat{\beta}_1)^2}{Q(\theta_1)} \right\}^{-\frac{n+a-r}{2}} \int_0^A \frac{dz}{(1+z^2)^{\frac{n+a-r+1}{2}}} . \end{aligned}$$

In order to see the propriety of the posterior, it is more convenient to look at $\pi(\boldsymbol{\beta}, \sigma | \mathbf{y})$. This is given by

$$\begin{aligned} \pi(\boldsymbol{\beta}, \sigma | \mathbf{y}) &\propto \sigma^{-(n+a)} |\beta_2| (c_{11}\beta_1^2 + 2c_{12}\beta_1\beta_2 + c_{22}\beta_2^2)^{-\frac{1}{2}} \\ &\quad \times \exp \left[-\frac{1}{2\sigma^2} \left\{ SSE + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^T \mathbf{X}^T \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right\}^{-\frac{n+a-1}{2}} \right] , \end{aligned}$$

Now, integrating out with respect to $\beta_3, \dots, \rho_r, \sigma$, one gets

$$\begin{aligned} \pi(\beta_1, \beta_2 | \mathbf{y}) &\propto |\beta_2| (c_{11}\beta_1^2 + 2c_{12}\beta_1\beta_2 + c_{22}\beta_2^2)^{-\frac{1}{2}} \\ &\quad \times \left\{ SSE + c_{11}(\beta_1 - \hat{\beta}_1)^2 + 2c_{12}(\beta_1 - \hat{\beta}_1)(\beta_2 - \hat{\beta}_2) + c_{22}(\beta_2 - \hat{\beta}_2)^2 \right\}^{-\frac{n+a-r+1}{2}} \end{aligned}$$

Recognizing the above pdf is proportional to $E(|\beta_2| | \mathbf{y})$ from a bivariate t -distribution with location parameters $\hat{\beta}_1$ and $\hat{\beta}_2$, scale matrix $\frac{1}{2}[SSE]\mathbf{C}^{-1}$ and degrees of freedom $\frac{n+a-r-1}{2}$, the propriety of the posterior follows when $n+a > r+3$.

CHAPTER 5 SUMMARY AND FUTURE RESEARCH

5.1 Summary

This dissertation develops “vague” or “noninformative” priors for several classical problems of statistical inference. In particular, we revisit the celebrated Behrens-Fisher problem, common mean estimation problem and problems involving certain regression ratios in multiple regression models. The latter includes as special cases (i) the Fieller-Creasy problem, (ii) the calibration problem, (iii) slope-ratio assays, and (iv) parallel-line assays.

For the Behrens-Fisher problem, we developed several reference priors, as well as probability matching priors. Jeffreys’ independent prior turned out to be the one-at-a-time reference prior, while a new prior was found which turned out to be a second order probability matching prior, HPD matching prior as well as a matching prior via inversion of conditional likelihood test statistics. These priors were all superior to Jeffreys’ general rule prior which is proportional to the positive square root of the determinant of the Fisher information matrix.

For the common mean estimation problem, we found reference priors, also we found a prior which was second order matching, HPD matching as well matching via inversion of conditional likelihood test statistics. This prior produces credible intervals which have very good frequentist coverage probabilities and is nearly equivalent to the optimal frequentist procedure in every situation considered. Incidentally, in this example, there does not exist any single “optimal” frequentist interval.

Chapter 4 developed priors suitable for inference involving ratios of regression coefficients. In the process, for the specific examples mentioned earlier, some of the

previously proposed priors were found as special cases, while some new priors were proposed.

5.2 Future Research

We want to consider some multivariate extensions of the proposed methods. While the quantile matching priors require real valued parameter of interest, the same is not required of HPD matching priors or matching via inversion of conditional likelihood ratio test statistics. Behrens-Fisher problem as well as the multivariate common mean estimation problem and develop the reference priors as well matching priors for the problem.

REFERENCES

- Aspin,A.A.(1948). An estimation and further development of a formula arising in the problem of comparing two mean values. *Biometrika*, **35**, 88-96.
- Bartlett,M.S.(1936). The information available in small samples. *Proc.Cambridge Philos.Soc.*, **32**, 560-566.
- Behrens,B.V.(1929).*Landwirtch.Jb.*, **68**, 807-837.
- Berger,J.O., and Bernardo,J.M.(1989). Estimating a product of means: Bayesian analysis with reference priors. *JASA*, **84**, 200-207.
- Berger,J.O., and Bernardo,J.M.(1992a). On the development of reference priors (with discussion). In J.M.Bernardo, J.O.Berger, A.P.Dawid & A.F.M.Smith, eds, *Bayesian Statistics 4*, 35-60, *Oxford University, Press*
- Berger,J.O., and Bernardo,J.M.(1992b). Ordered group reference priors with application to the multinomial problem *Biometrika*, **79**, 25-37.
- Berger,J.O., Liseo,B., and Wolpert,R.(1996). Integrated likelihood methods for eliminating nuisance parameters. *Tech.Report #96-7C B*, Dept.of Stat., Purdue University.
- Bernardo,J.M.(1977). Inference about the ratio of normalmeans: A Bayesian approach to the Fieller-Creasy problem. In *Recent Development in Statistics*, J.R.Barra et al. 345-350. Amsterdam,North-Holland.
- Bernardo,J.M.(1979). Reference posterior distributions for Bayesian inference. *JRSS B*, **41**, 113-147.
- Bhattacharya,C.G.(1980). Estimation of common mean and recovery of interblock information. *Ann. Statist.*, **8**, 205-211.
- Brown,L.D., and Cohen,A.(1974). Point and confidence estimation of a common mean and recovery of inter-block information. *Ann. Statistics*, **2**, 963-976.-69.
- Brown,P.J.(1982). Multivariate calibration (with discussion). *JRSS B*, **44**, 287-321.

- Buehler, R.J. (1959). Some validity criteria for statistical inference. *Ann. Math. Statistics*, **30**, 845-863.
- Buonaccorsi, J.P., and Gatsonis, C.A. (1988). Bayesian inference for ratios of coefficients in a linear model. *Biometrics*, **44**, 87-101.
- Cohen, A., and Sackrowitz, H.B. (1974). On estimating of the common mean of two populations. *Ann. Statistics*, **2**, 1274-1282.
- Cox, D.R., and Hinkley, D.V. (1974). *Theoretical Statistics*. Chapman & Hall, New York.
- Cox, D.R., and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *JRSS B*, **49**, 1-39.
- Creasy, M.A. (1954). Limits for the ratio of the means. *JRSS B*, **16**, 186-194.
- Datta, G.S. (1996). On priors providing frequentist validity of Bayesian inference for multiple parameter functions. *Biometrika*, **83**, 287-298.
- Datta, G.S., and Ghosh, J.K. (1995). On priors providing frequentist validity of Bayesian inference. *Biometrika*, **82**, 37-45.
- Datta, G.S., and Ghosh, M. (1995). Some remarks on noninformative priors. *JASA*, **90**, 1357-1363.
- Datta, G.S., and Ghosh, M. (1996). On the invariance of noninformative priors. *Ann. Stat.*, **24**, 141-159.
- Dharmadhikari, S., and Joag-dev, K. (1988). *Unimodality, Convexity and Applications*. Academic Press, New York.
- DiCiccio, T.J., and Stern, S.E. (1994). Frequentist and Bayesian Bartlett correction of test statistics based on adjusted profile likelihood. *JRSS B*, **56**, 397-408.
- Eberhardt, K.R., Reeve, C.P., and Spiegelman, C.H. (1989). A minimax approach to combining means, with practical examples. *Chemometrics and Intelligent Laboratory Systems*, **5**, 129-148.

- Fairweather, W.R. (1972). A method of obtaining an exact confidence interval for the common mean of several normal populations. *Applied Statistics*, **21**, 229-233.
- Filler, E.C. (1954). Some problems in interval estimation. *JRSS B*, **16**, 175-185.
- Finney, D.J. (1978) *Statistical Method in Biological Assay*. Charles Griffin & Company: London.
- Fisher, R.A. (1932). *Statistical Methods for Research Workers*. Oliver & Boyd: London.
- Fisher, R.A. (1935). The fiducial argument in statistical inference. *Ann. Eugenics*, **11**, 141-172.
- Fisher, R.A. (1956). On a test of significance in Pearson's Biometrika Tables (No. 11). *JRSS B*, **18**, 56-60.
- Garvan, C.W., and Ghosh, M. (1997). Noninformative priors for dispersion models. *Biometrika*, **84**, 4, 976-982.
- George, E.O. (1977). Combining independent one-sided and two-side statistical tests—some theory and applications. Unpublished Ph.D. Dissertation, University of Rochester.
- Ghosh, J.K. (1994). Higher Order Asymptotics. *NSF-CBMS Regional Conference Series in Probability and Statistics*, **4**, Institute of Mathematical Statistics.
- Ghosh, J.K., and Mukerjee, R. (1991). Characterization of priors under which Bayesian and frequentist Bartlett corrections are equivalent in the multiparameter case. *J. Mult. Anal.*, **38**, 385-393.
- Ghosh, J.K., and Mukerjee, R. (1992). Bayesian and frequentist Bartlett corrections for likelihood ratio and conditional likelihood ratio test. *JRSS B*, **54**, 867-875.
- Ghosh, J.K., and Mukerjee, R. (1995). Frequentist validity of highest posterior density regions in the presence of nuisance parameters. *Statistics & Decisions*, **13**, 131-139.

- Ghosh,M., Carkin,B.P., and Srivastava,M.S.(1995). Probability matching priors for linear calibration. *Test*, **4**, No.2, 333-357.
- Ghosh,M., and Yang,M.M.(1996). Noninformative priors for the two sample normal problem. *Test*, **5**, 145-157.
- Ghosh,M., and Yin,Min.(2000). *Empirical Bayes and Likelihood Inference*. Springer Verlag: New York.
- Gleser,L.J., and Hwang,J.T.(1987). The non-existence of 100 $(1 - \alpha\%)$ confidence sets of finite expected diameters in error-in-variable and related models. *Ann. Stat.*, **15**, 1351-1362.
- Graybill,F.A., and Deal,R.B.(1959). Combining unbiased estimators. *Biometrics*, **15**, 543-550.
- Hedges,L.V., and Oklin,I.(1985). *Statistical Methods for Meta-Analysis*. Academic Press: New York.
- Hoadley,B.(1970). A Bayesian lookat inverse linear regression. *JASA*, **65**, 356-369.
- Hill,B.M.(1981). Discussion of "A Bayesian analysis of the linear calibration problem". *Technometrics*, **23**, 335-338.
- Hunter,W.G., and Lamboy,W.F.(1981). A Bayesian analysis of the linear calibration problem. *Technometrics*, **23**, 323-350.
- Jeffreys,H.(1961). *Theory of probability*. Oxford University Press: London.
- Jordan,S.M., and Krishnamoorthy,K.(1996). Exact confidence intervals for the common mean of several normal populations. *Biometrics*, **52**, 77-86.
- Jørgensen,B.(1992). *The Theory of Exponential Dispersion Models and Analysis of Deviance*. Rio de Janeiro : Conselho Nacional de Desenvolvimento Cientifico e Tecnologico, Instituto de Matematica Pura e Aplicada.
- Kappenman,R.F, Geisser,S., and Antle,C.F.(1970). Bayesian and fiducial solutions to the Fieller-Creasy problem. *Sankhya B*, **32**, 331-340.

- Kass, R.E., and Wasserman, L.(1996). Formal rules for selecting prior distributions: a review and annotated bibliography. *JASA*, **91**, 1343-1370.
- Kendall,M.G., and Stuart,A.(1967). *The advanced Theory of Statistics*, **2**,: Inference and Relationship, 2nd ed. Charles Griffin: London.
- Kharti,C.G., and Shah,K.R.(1974). Estimation of location parameters from two linear models under normality. *Comm. Statist. Theory and Methods*, **3**, 647-663.
- Kim,P.T., Carter,E.M., and Hubert,J.J.(1991) . Estimating relative potency using prior information. *Biometrics*, **47**, 295-301.
- Krutchkoff,R.G.(1967). Classical and inverse regression methods of calibration. *Technometrics*, **9**, 429-439.
- Kubokawa,T.(1990). Minimax estimation of common coefficient of several regression models under quadratic loss. *J. Statist. Plan. and Inf.*, **24**, 337-345
- Laplace,P.(1812). *Theorie Analytique des Probabilities*. Courcier: Paris.
- Lawless,J.F.(1981). Discussion of "A Bayesian analysis of the linear calibration problem". *Technometrics*, **23**, 334-335.
- Lee,A., and Gurland,J.(1975). Size and power of tests for equality of means of two normal populations with unequal variances. *JASA*, **70**, 933-947.
- Lindley,D.V.(1958). Fiducial distributions and Bayes' theorem. *JRSS B*, **20**, 102-107.
- Liseo,B.(1993). Elimination of nuisance parameters with reference priors. *Biometrika*, **80**, 295-304.
- Meier,P.(1953). Variance of weighted mean. *Biometrics*, **9**, 59-73.
- Mendoza,M.(1988). Inference about the ratio of linear combinations of the coefficients in a multiple regression model. *Bayesian Statistics 3* (J.M.Bernardo, M.H.DeGroot, D.V.Lindley and A.F.M.Smith. eds). Oxford University Press: London. 705-711.

- Mendoza,M.(1990). A Bayesian analysis of the slope ratio bioassay. *Biometrics*, **46**, 1059-1069.
- Mendoza,M.(1996). A note on the confidence probabilities of reference prior for the calibration model. Preprint.
- Mukerjee,R., and Dey,D.K.(1993). Frequentist validity of posterior quantiles in the presence of a nuisance parameter: higher order asymptotics. *Biometrika*, **80**, 499-505.
- Mukerjee,R., and Ghosh,M.(1997). Second order probability matching priors. *Biometrika* , **84**, 970-975.
- Neyman,J., and Scott,E.L.(1948). Consistent estimates based on partially consistent observations. *Econometrika*, **16**, 1-32.
- Nicolaou,A.(1993). Bayesian intervals with good frequentist behavior in the presence of nuisance parameters. *JRSS B*, **27**, 9-16.
- Peers,H.W.(1965). On confidence sets and Bayesian probability points in the case of several parameters. *JRSS B*, **27**, 9-16.
- Philippe,A., and Robert,C.(1998). A note on the confidence properties of reference priors for the calibration model. *Test*, **7**, 147-160.
- Reid, N.(1995). Likelihood and Bayesian approximation methods. In *Bayesian Statistics 5*. Eds. J.M. Bernardo, J. Berger, A.P. Dawid and A.F.M.Smith. Oxford University Press, pp 351-368.
- Robinson,G.K.(1976). Properties of Student's t and of the Behrens-Fisher solution to the two means problem. *Annals of Statistics*, **4**, 963-971.
- Robinson,G.K.(1982). Behrens-Fisher problem. In *Encyclopedia of Statistical Sciences*, V1. Eds. N.L. Johnson, S. Kotz and C.B. Read. Wiley, New York, pp 205-209.
- Scheffé,H.(1970). Practical solutions of the Behrens-Fisher problem. *JASA*, **65**, 1501-1508.

- Sendra,M.(1982). Distribucion final de referencia para el problema de Fieller-Creasy. *Trabajos de Estadistica Investigation Operation*, **33**, 55-72.
- Severini,T.(1991). On the relationship between Bayesian and non-Bayesian interval estimates. *JRSS B*, **53**, 611-618.
- Shinozaki,N.(1978). A note on estimating the common mean of k normal distributions and the Stein problem. *Comm. Statist. Theory Methods*, **7**, 1421-1432.
- Sinha,B.K.(1985). Unbiased estimation of the variance of the Graybill-Deal estimator of the common mean of several normal populations. *Canadian J.of Statistics*, **13**, 243-247.
- Sinha,B.K.(1998). *Statistical Meta-Analysis with Applications*.
- Sinha,B.K., and Mouqadem,O.(1982). Estimation of the common mean of two univariate normal populations. *Comm. Statist. Theory and Methods*, **11**, 1603-1614.
- Snedecor,G.W.(1950). The statistical part of the scientific method. *Annals of the New York Academy of Science*, **52**, 742-749.
- Stein,C.(1985). On coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics*, **16**, 485-514. Banach Center Publications: Warsaw.
- Stephens,D.A., and Smith,A.F.M.(1992). Sampling-resampling techniques for the computation of posterior densities in normal means problem. *Test*, **1**, 1-18.
- Stouffer,S.A., Suchman,E.A., DeVinney,L.C. Star,S.A. and Williams, R.M.Jr. (1949). *The American Soldier*, **1** Princeton, NJ: Princeton University Press.
- Sun,D., and Ye,K.(1996). Frequentist validity of posterior quantiles for a two parameter exponential family. *Biometrika*, **83**, 55-65.
- Tibshirani,R.(1989). Noninformative priors for one parameter of many. *Biometrika*, **76**, 604-608.
- Tippett,L.H.(1931). *The Methods of Statistics*. London: Williams & Norgate.

- Welch, B.L. (1947). The generalization of Student's problem when several different population variances are involved. *Biometrika*, **34**, 28-35.
- Welch, B.L., and Peers, H.W. (1963). On formula for confidence points based on integrals of weighted likelihoods. *JRSS B*, **25**, 318-329.
- Wood, E.C. (1946). The theory of certain analytical procedures, with particular reference to microbiological assays. *Analyst*, **71**, 1-14.
- Yin, M., and Ghosh, M. (1997). A note on the probability difference between matching priors based on posterior quantiles and on inversion of conditional likelihood ratio statistics. *Calcutta Statistical Association Bulletin*, **47**, 185-186.
- Yu, P.L., Sun, Y., and Sinha, B.K. (1999). On exact confidence intervals for the common mean of several normal populations. *J. Statist. Plan. and Inf.* To Appear 2000.
- Zacks, S. (1966). Unbiased estimation of the common mean. *JASA*, **61**, 467-476.
- Zacks, S. (1970). Bayes and fiducial equivariant estimators of the common means of two populations. *Ann. Math. Statistics*, **41**, 59-69.

BIOGRAPHICAL SKETCH

Yeong-Hwa Kim was born on April 15, 1964, in Seoul, Korea. In 1990, he earned a Bachelor of Economics in Statistics degree from Chung-Ang University, Seoul, Korea, and two years later earned a Master of Economics in Statistics degree in February, 1992, from the same university. He entered the PhD program in Statistics at the same university in March 1992. With highly motivated reasons, however, he quit his PhD work in Korea on August, 1994. Finally, he entered the PhD program in Statistics at the University of Florida in August, 1995. He served part-time as an instructor in many universities in Korea from 1992 until 1995. In addition to pursuing his Ph.D. in Statistics from the University of Florida, he has served as a teaching assistant of the Department of Statistics at UF.

He got married in 1990 and is the father of a wonderful daughter and a son. After graduation, the author will begin his next adventure as a researcher at the Research Institute of Finance of Samsung Life Insurance Co. in Seoul, Korea.

NONINFORMATIVE PRIORS AND BAYESIAN INFERENCE

Yeong-Hwa Kim

(352) 338-0858

Department of Statistics

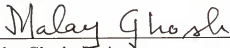
Chair: Dr. Malay Ghosh

Degree: Doctor of Philosophy

Graduation Date: August 2000

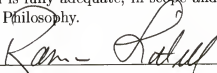
In recent years, statistics has been widely used in most academic and industrial fields. In addition to the information obtained from a sample Bayesian considers prior information of the parameters. Naturally, the first step of Bayesian inference is to find an appropriate prior distribution. Bayesian analysis is performed by combining the prior information and the sample information into what is called the posterior distribution. Just as the prior distribution reflects beliefs prior to the experimentation, so the posterior reflects the updated beliefs after observing the sample. In other words, the posterior distribution combines the prior information about the parameter of interest contained in the sample to give a composite picture of the final beliefs about the parameter of interest. Bayesian techniques have found wider acceptance in recent years in the theory and practice of statistics. These can be explained by the fact that even in the presence of little or vague prior information, Bayesian techniques can often be used successfully by employing the so-called "default" priors. Thus, not surprisingly, over the years, a wide range of noninformative priors has been proposed and studied. This dissertation focused on Bayesian analysis of some classical problems of statistical inference. In particular, we have derived several "default" priors such as Jeffreys' prior, reference priors, and probability matching priors for the Behrens-Fisher problem, for the common mean of several normal populations, and for the ratios of regression coefficients in linear models. The latter includes as special cases, the linear calibration problem, the Fieller-Creasy problem, slope-ratio bioassay, and parallel-line bioassay.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Malay Ghosh, Chairman
Distinguished Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



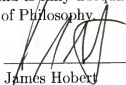
Ramon Littell
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



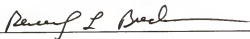
James Booth
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



James Hoberg
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Beverly Brechner
Professor of Mathematics

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

August 2000

Dean, Graduate School